JOURNAL OF COMPUTER AND SYSTEM SCIENCES 44, 220–258 (1992)

Almost Everywhere High Nonuniform Complexity*

JACK H. LUTZ

Department of Computer Science, Iowa State University, 226 Atanasoff Hall, Ames, Iowa 50011

Received August 1, 1989; revised March 8, 1991

We investigate the *distribution* of nonuniform complexities in uniform complexity classes. We prove that almost every problem decidable in exponential space has essentially maximum circuit-size and space-bounded Kolmogorov complexity almost everywhere. (The circuit-size lower bound actually exceeds, and thereby strengthens, the Shannon $2^n/n$ lower bound for almost every problem, with no computability constraint.) In exponential time complexity classes, we prove that the strongest relativizable lower bounds hold almost everywhere for almost all problems. Finally, we show that infinite pseudorandom sequences have high nonuniform complexity almost everywhere. The results are unified by a new, more powerful formulation of the underlying measure theory, based on uniform systems of density functions, and by the introduction of a new nonuniform complexity measure, the *selective* Kolmogorov complexity.

1. INTRODUCTION

A precise account of the quantitative relationships between uniform and nonuniform complexity measures is a principal objective of the theory of computation. For the most important nonuniform complexity measures—those that measure size of programs and size of circuits—this paper establishes new lower bounds that hold almost everywhere in uniform time and space complexity classes.

The circuit-size complexity of Boolean functions has been studied for over 50 years. Shannon [38] proved that every Boolean function $f: \{0,1\}^n \to \{0,1\}$ is computed by a circuit with $O(2^n/n)$ gates and that, asymptotically, almost every such function requires more than $2^n/n(1-\varepsilon)$ gates, for every $\varepsilon > 0$. Lupanov [23] tightened Shannon's upper bound by proving that every such function f is computed by a circuit with $(2^n/n)(1+O(1/\sqrt{n}))$ gates. Since Lupanov's upper bound and Shannon's lower bound have asymptotic ratio 1, these bounds together imply that almost every Boolean function has essentially maximum circuit-size complexity. Lupanov named this phenomenon the *Shannon effect*.

In order to compare circuit size to uniform, algorithmic complexity measures, the circuit-size complexity measure has been extended in the natural way from Boolean

* This research was supported in part by NSF Grants CCR-8809238 and CCR-9157382 and in part by DIMACS, where the author was a visitor while part of this work was performed.

0022-0000/92 $3.00

functions to *decision problems*, i.e., to (infinite) binary sequences $x \in \{0,1\}^\infty$. In this setting, a routine modification of Shannon's lower bound argument gives the following formulation of the Shannon effect. If $\varepsilon > 0$ and an infinite binary sequence is chosen probabilistically by using an independent toss of a fair coin to decide each bit, then with probability 1 the chosen sequence x will have circuit-size complexity $CS_x(n) > (2^n/n)(1-\varepsilon)$ for all but finitely many n. More succinctly, in the usual Lebesgue measure on $\{0,1\}^\infty$, almost every binary sequence x has $CS_x(n) > (2^n/n)(1-\varepsilon)$ for almost every n.

The set P/Poly, consisting of those decision problems that have polynomial-size circuits, is of particular interest. It is clear that P/Poly is an uncountable, measure 0 subset of $\{0,1\}^\infty$ and that $P \subseteq P/Poly$. Kannan [15] has shown that $ESPACE \not\subseteq P/Poly$. It is widely believed that $NP \not\subseteq P/Poly$, i.e., that NP-complete problems are infeasible in a strong, information-theoretic sense. Supporting this conjecture, Karp and Lipton [16] have shown that $NP \subseteq P/Poly$ has the unlikely consequence of collapsing the polynomial-time hierarchy to its second level. On the other hand, Wilson [44] has exhibited oracles relative to which $E_2 = DTIME(2^{poly}) \subseteq P/Poly$ and problems in NP and $E = DTIME(2^{linear})$ all have *linear*-size circuits, so progress toward resolving this conjecture may not come easily.

A *distributional* investigation of uniform versus nonuniform complexity was initiated by Lutz [24]. Regarding the Kannan $ESPACE \not\subseteq P/Poly$ result, we addressed the following question. Among problems in ESPACE, is the phenomenon of not having polynomial-size circuits rare, or is it in some sense typical? This question led to the development of *resource-bounded category and measure* in [24]. These techniques, which extend classical and effective versions of Baire category and Lebesgue measure (see [33, 9, 7, 30, 31]), define the *meager* ("topologically small") and *measure 0* ("probabilistically small") subsets of various complexity classes, respectively. It was proven in [24] that $P/Poly \cap ESPACE$ is a meager, measure 0 subset of ESPACE. Thus the phenomenon of not having polynomial-size circuits is *very* typical of problems in ESPACE, in the sense of both category and measure.

In this paper we prove that the Shannon effect holds with full force in ESPACE. Specifically, with respect to measure, for every real $\alpha < 1$, almost every binary sequence $x \in ESPACE$ has circuit-size complexity $CS_x(n) > (2^n/n)(1 + \alpha \log n/n)$ for almost every n. This almost everywhere lower bound on circuit-size complexity in ESPACE extends the previous work in two significant ways.

(i) The $(2^n/n)(1 + \alpha \log n/n)$ lower bound here exceeds the $o(2^n/n)$ lower bound of [24] and is only negligibly smaller than the Lupanov $(2^n/n)(1 + O(1/\sqrt{n}))$ upper bound for every $x \in \{0,1\}^\infty$. (In fact, the present lower bound slightly exceeds, and as a consequence tightens, the Shannon $2^n/n$ lower bound for almost every $x \in \{0,1\}^\infty$.)

(ii) The lower bound here is proven to hold for *almost every n*, whereas the lower bound in [24] is only shown to hold for *infinitely many n*. For example, let

$P/Poly^{i.o.}$ be the set of binary sequences x for which there is a polynomial q such that $CS_x(n) \leqslant q(n)$ for infinitely many n. The proof of Kannan [15] actually shows that $ESPACE \not\subseteq P/Poly^{i.o.}$. The present result implies that $P/Poly^{i.o.} \cap ESPACE$ is in fact a measure 0 subset of ESPACE.

Putting these advances together gives our strong formulation of the Shannon effect in ESPACE: *almost every* problem in ESPACE has *essentially maximum* circuit-size complexity *almost everywhere.*

The Kolmogorov complexity (often called the program-size complexity) of binary strings and sequences was discovered independently by Solomonoff [40], Kolmogorov [18], and Chaitin [6]. The extraordinary power and scope of this notion have recently been surveyed by Kolmogorov and Uspenskii [19] and Li and Vitanyi [21]. In this paper we are primarily concerned with *resource-bounded* Kolmogorov complexities, which have been investigated by Hartmanis [10], Sipser [39], Ko [17], Longpré [22], Balcázar and Book [3], Huynh [13], Lutz [24], Allender and Watanabe [2], and many others.

Martin-Löf [29] showed that $K(x|n)$, the conditional Kolmogorov complexity of infinite binary sequences x, exhibits a strong Shannon effect. Specifically, Martin-Löf proved that if the series $\sum_{n=0}^{\infty} 2^{-f(n)}$ converges (e.g., if $f(n) = \alpha \log n$ for some real $\alpha > 1$), then in the sense of Lebesgue measure, almost every binary sequence $x \in \{0, 1\}^{\infty}$ has conditional Kolmogorov complexity $K(x|n) > n - f(n)$ for all but finitely many n. For suitable f, this lower bound is already very close to the well-known upper bound, $K(x|n) < n + c$ for all x and n, where c is a fixed constant. However, Martin-Löf [29] also tightened the upper bound by proving that if f is computable and the series $\sum_{n=0}^{\infty} 2^{-f(n)}$ diverges (e.g., if $f(n) = \log n$), then every binary sequence $x \in \{0, 1\}^{\infty}$ has conditional Kolmogorov complexity $K(x|n) < n - f(n)$ for infinitely many n. Thus, for computable f, it is the convergence/divergence behavior of $\sum_{n=0}^{\infty} 2^{-f(n)}$ that determines whether $n - f(n)$ is an infinitely often upper bound on $K(x|n)$ for all x or an almost everywhere lower bound on $K(x|n)$ for almost every x. Since the convergence/divergence behavior of $\sum_{n=0}^{\infty} 2^{-f(n)}$ is sensitive to very small changes in the growth rate of f, this implies that almost every binary sequence $x \in \{0, 1\}^{\infty}$ has essentially maximum conditional Kolmogorov complexity almost everywhere.

We prove in this paper that the Shannon effect holds with full force, in essentially the above form, for the space-bounded conditional Kolmogorov complexity of problems in ESPACE. Moreover, we unify this result with the Shannon effect for circuit size in ESPACE by introducing a new program-size complexity measure, the *selective* Kolmogorov complexity. Roughly speaking, the conditional Kolmogorov complexity of x at n, written $K(x|n)$, is the length of the shortest program π that, given n, outputs the first n bits of x. The σ-selective Kolmogorov complexity of x at n, written $K(x \wedge \sigma|n)$, is the same, except that the program π is now only required to be correct about bits of x specified by $\sigma(n)$, the value of the *selector* σ at n. If the selector σ requires all bits to be correct, then $K(x \wedge \sigma|n) = K(x|n)$; i.e., the σ-selective Kolmogorov complexity is precisely the conditional Kolmogorov

complexity. However, if $\sigma(n)$ only requires π to be correct about some of the first n bits of x, then $K(x \wedge \sigma \mid n)$ may be much smaller than $K(x \mid n)$.

The main theorem of this paper is Theorem 4.4, which shows that almost every problem in ESPACE has very high space-bounded selective Kolmogorov complexity almost everywhere. By inequality (4.4) this almost everywhere lower bound is tight, so we have a strong instance of the Shannon effect: *almost every* problem in ESPACE has *essentially maximum* space-bounded selective Kolmogorov complexity *almost everywhere.*

This appears to be a very powerful formulation of the Shannon effect in ESPACE. The above-mentioned Shannon effects for circuit-size and conditional Kolmogorov complexities in ESPACE are derived from this more general result.

We also prove almost everywhere lower bounds for nonuniform complexities in uniform time complexity classes. In this case our lower bounds are considerably smaller than known upper bounds, so much remains to be discovered. From a *distributional* point of view, however, our results are quite strong. We prove that the highest levels of circuit-size and time-bounded Kolmogorov complexity known (or provable by relativizable methods) to be exceeded infinitely often by *any* problem decidable in exponential time are in fact exceeded *almost everywhere* by *almost every* problem decidable in exponential time.

Our almost everywhere lower bounds on nonuniform complexity have immediate consequences for the theory of pseudorandom sequences. Following work by Yao [45], Blum and Micali [5], Goldreich, Goldwasser, and Micali [8], Levin [20], Allender [1], and others on the generation of finite pseudorandom sequences from shorter random sequences, and following work by Schnorr [34, 36], Wilber [43], Huynh [12, 13], Ko [17], and others on pseudorandom properties of infinite sequences, Lutz [25, 27] gave a measure-theoretic definition of infinite pseudorandom sequences. This definition of pseudorandomness is analogous to the Martin-Löf [28] definition of randomness, but is based on resource-bounded measure theory and thereby provides an abundance of pseudorandom sequences that are deterministically computable at relatively low complexity levels. Pseudorandom sequences and their properties are discussed in detail in [27]. In this paper we use our almost everywhere lower bounds to show that infinite pseudorandom sequences have high circuit-size and Kolmogorov complexity almost everywhere.

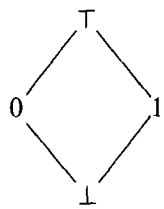
Note. The resource-bounded measure introduced in [24], and used to define pseudorandom sequences in [25], was formulated in terms of "covering by modulated enumerations of cylinders." This formulation is not strong enough (i.e., does not render enough sets measurable) to prove the main results of the present paper. Indeed, some of the proofs in [24, 25] are not correct without some technical modification of the underlying measure theory. In Section 3 below, we present some of the elements of a new, more powerful formulation of resource-bounded measure, based on *uniform systems of density functions.* This formulation, like the old one, is a general theory with a *resource bound* (class of functions) Δ as a parameter. Various choices of this parameter Δ give various measure theories as

special cases. One of these cases is classical Lebesgue measure. Other special cases impose internal measure-theoretic structure on REC, E, ESPACE, and other complexity classes. All sets measurable in the formulation of [24] have the same measure in the new formulation, and the new formulation admits rigorous (and simpler) proofs of the applications in [24, 25]. Moreover, the new formulation, by expanding the class of measurable sets, has yielded a number of new applications, especially in time-bounded complexity classes.

Although a complete development of resource-bounded measure is beyond the scope of this paper, Section 3 below presents all the ideas, results, and proofs needed for the applications here. The present paper is thus self-contained. (Theorems 3.19, 6.2, and 6.3 are not proven or used in this paper.) A brief discussion of the relation between our density functions and the martingales used by Schnorr [34–37] in his investigation of random and pseudorandom sequences appears at the end of Section 3. More thorough discussions of resource-bounded measure and pseudorandomness will appear in [26, 27].

2. Preliminaries

We work in two alphabets, the usual binary alphabet $\{0, 1\}$ and the extended binary alphabet $\Sigma = \{0, 1, \bot, \top\}$. The elements $\bot$ ("bottom") and $\top$ ("top") of Σ are interpreted as "undefined bit" and "impossibly defined bit," respectively. We define $\sqsubseteq$ to be the partial ordering



of Σ. Thus $b \sqsubseteq b'$ means that bit b is "no more defined than" bit b'.

A *string* is a finite sequence $x \in \Sigma^*$. A *binary string* is a string $x \in \{0, 1\}^*$. A *sequence* is an infinite sequence $x \in \Sigma^\infty$. A *binary sequence* is a sequence $x \in \{0, 1\}^\infty$. We use variables x, y, z, etc., to denote strings or sequences. We write $|x|$ for the length of x. Thus $|x| \in \mathbf{N} \cup \{\infty\}$, where $\mathbf{N}$ is the set of nonnegative integers. The unique string of length 0 is λ, the empty string.

If x is a string and y is a string or sequence, then xy is the concatenation of x and y. If x is already a sequence, then $xy = x$. If x is a string and $k \in \mathbf{N} \cup \{\infty\}$, then x^k is the k-fold concatenation of x with itself. Thus $x^0 = \lambda$ and $x^{k+1} = xx^k$.

If $0 \leqslant i \leqslant j < |x|$, then $x[i..j]$ is the string consisting of the ith through jth bits of x. Thus $x = x[0..|x| - 1]$ if x is a string. We write $x[i]$ for $x[i..i]$, the ith bit of x.

We extend the partial ordering $\sqsubseteq$ to strings and sequences via the following rules.

(i) For $x, y \in \Sigma^\infty$, $x \sqsubseteq y$ if and only if $x[i] \sqsubseteq y[i]$ for every $i \in \mathbf{N}$.

(ii) For $x, y \in \Sigma^\infty$, $x \sqsubsetneq y$ if and only if $x \sqsubseteq y$ and $x \neq y$.

(iii) For arbitrary x and y, $x \sqsubseteq y$ if and only if $x\perp^\infty \sqsubseteq y\perp^\infty$.

(iv) For arbitrary x and y, $x \sqsubsetneq y$ if and only if $x\perp^\infty \sqsubsetneq y\perp^\infty$.

The extended relation $\sqsubseteq$ is not technically a partial ordering because it is not antisymmetric. For example, for any string x, x and $x\perp$ are distinct strings with $x \sqsubseteq x\perp$ and $x\perp \sqsubseteq x$. In practice, however, we will think of x, $x\perp$, and $x\perp^\infty$ as denoting essentially the same object, so no confusion will result from calling $\sqsubseteq$ a partial ordering of strings and sequences. Note that $x \sqsubsetneq y$ means that x is "strictly less defined than" y. Thus, for example, it is *not* the case that $x \sqsubsetneq x\perp$.

Note that if x and y are binary strings, i.e., $x, y \in \{0, 1\}^*$, then $x \sqsubseteq y$ means that x is a prefix of y and $x \sqsubsetneq y$ means that x is a proper prefix of y.

We define $\|x\|$, the number of defined bits in a string $x \in \Sigma^*$, by the following recursion.

$$\|\lambda\| = 0$$

$$\|x\perp\| = \|x\|$$

$$\|x0\| = \|x1\| = \|x\| + 1$$

$$\|x\top\| = \infty.$$

Thus $\|x\| \leqslant |x|$ if $x \in \{0, 1, \perp\}^*$, $\|x\| = |x|$ if $x \in \{0, 1\}^*$, and $\|x\| = \infty$ if x contains any occurrence of $\top$.

Our primary objects of study are the binary sequences. The extended binary alphabet Σ is a technical device whose primary role is the following.

DEFINITION 2.1. The *cylinder generated by* a string $x \in \Sigma^*$ is

$$C_x = \{y \in \{0, 1\}^\infty \mid x \sqsubseteq y\}.$$

Thus we regard a string $x \in \Sigma^*$ as an approximation, or "partial specification" of a binary sequence y. The cylinder C_x is the set of all binary sequences that meet this specification. If $\top$ appears in x, then $C_x = \varnothing$; i.e., the specification x is unsatisfiable.

The *measure* of a cylinder C_x is $\mu(x) = \mu(C_x) = 2^{-\|x\|}$. This is the probability that $y \in C_x$ when the binary sequence $y \in \{0, 1\}^\infty$ is chosen probabilistically by using an independent toss of a fair coin to decide each bit of y.

It is useful to have an operation that "merges" two specifications. To this end, for $b, b' \in \Sigma$, we write $b \wedge b'$ for the least upper bound of b and b' with respect to $\sqsubseteq$. We then extend the operation $\wedge$ to strings and sequences as follows.

(v) For $x, y \in \Sigma^\infty$, $x \wedge y \in \Sigma^\infty$ is defined by $(x \wedge y)[i] = x[i] \wedge y[i]$ for all $i \in \mathbf{N}$.

(vi) For arbitrary x and y, $|x \wedge y| = \max\{|x|, |y|\}$ and $(x \wedge y)\perp^\infty = (x\perp^\infty) \wedge (y\perp^\infty)$.

It is easy to check that $\wedge$ does indeed merge specifications in the following sense.

Fact 2.2. For all $x, y \in \Sigma^*$, $C_{x \wedge y} = C_x \cap C_y$.

Complexity classes are usually defined as sets of languages. A *language* here is a set $L \subseteq \{0,1\}^*$, i.e., a set of binary strings. We fix the lexicographic enumeration $s_0 = \lambda$, $s_1 = 0$, $s_2 = 1$, $s_3 = 00, \ldots$ of $\{0,1\}^*$ and identify each language L with its *characteristic sequence* $x_L \in \{0,1\}^\infty$ defined by

$$x_L[k] = \begin{cases} 1 & \text{if } s_k \in L \\ 0 & \text{if } s_k \notin L. \end{cases}$$

This identifies the set $\mathscr{P}(\{0,1\}^*)$ of all languages with the set $\{0,1\}^\infty$ of all binary sequences. Under this identification, a string $x \in \Sigma^*$ *approximates* a language L, and we write $x \sqsubseteq L$, if $x \sqsubseteq x_L$. Thus the cylinder generated by x is also a set of languages,

$$C_x = \{L \subseteq \{0,1\}^* \mid x \sqsubseteq L\}.$$

We use X, Y, Z, etc., to denote sets of languages (equivalently, to denote sets of binary sequences). The *complement* of a set X is $X^c = \mathscr{P}(\{0,1\}^*) \setminus X = \{0,1\}^\infty \setminus X$.

We use the lexicographic successor function next: $\{0,1\}^* \to \{0,1\}^*$ defined by $\text{next}(s_k) = s_{k+1}$ for all $k \in \mathbf{N}$.

We fix once and for all a one-to-one pairing function $\langle\ ,\ \rangle$ from $\Sigma^* \times \Sigma^*$ onto Σ^* such that the pairing function and its associated projections, $\langle x, y\rangle \mapsto x$ and $\langle x, y\rangle \mapsto y$ are computable in polynomial time. We insist further that this pairing function satisfy the following conditions for all $x, y \in \Sigma^*$.

(a) $\langle x, y\rangle \in \{0,1\}^*$ if and only if $x, y \in \{0,1\}^*$.

(b) $\langle x, y\rangle \in \{0\}^*$ if and only if $x, y \in \{0\}^*$.

These conditions canonically induce pairing functions $\langle\ ,\ \rangle$ from $\{0,1\}^* \times \{0,1\}^*$ onto $\{0,1\}^*$ and from $\mathbf{N} \times \mathbf{N}$ onto $\mathbf{N}$, respectively. We write $\langle x, y, z\rangle$ for $\langle x, \langle y, z\rangle\rangle$, etc., so that tuples of any fixed length are coded by the pairing function.

We let $\mathbf{D} = \{m2^{-n} \mid m, n \in \mathbf{N}\}$ be the set of *nonnegative dyadic rationals*. Many functions in this paper take their values in $\mathbf{D}$ or in $[0, \infty)$, the set of nonnegative real numbers. In fact, with the exception of some functions that map into $[0, \infty)$, all our functions are of the form $f: X \to Y$, where each of the sets X, Y is $\mathbf{N}$, $\{0,1\}^*$, Σ^*, $\mathbf{D}$, or some cartesian product of these sets. Formally, in order to have uniform criteria for their computational complexity, we regard all such functions as mapping Σ^* into Σ^*. For example, a function $f: \mathbf{N}^2 \times \{0,1\}^* \to \mathbf{N} \times \mathbf{D}$ is formally interpreted as a function $\tilde{f}: \Sigma^* \to \Sigma^*$. Under this interpretation, $f(i, j, w) = (k, q)$ means that $\tilde{f}(\langle 0^i, \langle 0^j, w\rangle\rangle) = \langle 0^k, \langle u, v\rangle\rangle$, where u and v are the binary representations of the integer and fractional parts of q, respectively. Moreover, we only care

about the values of $\tilde{f}$ for arguments of the form $\langle 0^i, \langle 0^j, w\rangle\rangle$, and we insist that these values have the form $\langle 0^k, \langle u, v\rangle\rangle$ for such arguments.

For a function $f: \mathbf{N}\times X\to Y$ and $k\in\mathbf{N}$, we define the function $f_k: X\to Y$ by $f_k(x)=f(k,x)=f(\langle 0^k, x\rangle)$. We then regard f as a "uniform enumeration" of the functions $f_0, f_1, f_2, \ldots$. For a function $f: \mathbf{N}^n\times X\to Y$ $(n\geqslant 2)$, we write $f_{k,l}=(f_k)_l$, etc. For a function $f: \{0,1\}^*\to\{0,1\}^*$, we write f^n for the n-fold composition of f with itself.

We say that a condition $\Theta(n)$ holds *almost everywhere* (*a.e.*) if it holds for all but finitely many $n\in\mathbf{N}$. We say that $\Theta(n)$ holds *infinitely often* (*i.o.*) if it holds for infinitely many $n\in\mathbf{N}$.

We use the discrete logarithm

$$\log n=\min\{k\in\mathbf{N}\mid 2^k\geqslant n\}.$$

Note that $\log 0=0$.

For each $i\in\mathbf{N}$ we define a class G_i of functions from $\mathbf{N}$ into $\mathbf{N}$ as follows.

$$G_0=\{f\mid (\exists k)\, f(n)\leqslant kn \text{ a.e.}\}$$

$$G_{i+1}=2^{G_i(\log n)}=\{f\mid (\exists g\in G_i)\, f(n)\leqslant 2^{g(\log n)} \text{ a.e.}\}.$$

We also define the functions $\hat{g}_i\in G_i$ by $\hat{g}_0(n)=2n$, $\hat{g}_{i+1}(n)=2^{\hat{g}_i(\log n)}$. We regard the functions in these classes as growth rates. In particular, G_0 contains the linearly bounded growth rates and G_1 contains the polynomially bounded growth rates. It is easy to show that each G_i is closed under composition, that each $f\in G_i$ is $o(\hat{g}_{i+1})$, and that each $\hat{g}_i$ is $o(2^n)$. Thus G_i contains superpolynomial growth rates for all $i>1$, but all growth rates in the G_i-hierarchy are subexponential.

All results in this paper are robust with respect to reasonable choices of the underlying model of deterministic, algorithmic computation. Our *machines* and *algorithms* can thus be interpreted as Turing machines, random access machines, pointer machines, etc.

Within the class REC of all decidable languages, we are interested in the uniform complexity classes $\mathrm{E}_i=\mathrm{DTIME}(2^{G_{i-1}})$ and $\mathrm{E}_i\mathrm{SPACE}=\mathrm{DSPACE}(2^{G_{i-1}})$ for $i\geqslant 1$. The well-known exponential complexity classes $\mathrm{E}=\mathrm{E}_1=\mathrm{DTIME}(2^{\mathrm{linear}})$, $\mathrm{E}_2=\mathrm{DTIME}(2^{\mathrm{polynomial}})$, $\mathrm{ESPACE}=\mathrm{E}_1\mathrm{SPACE}=\mathrm{DSPACE}(2^{\mathrm{linear}})$, and $\mathrm{E}_2\mathrm{SPACE}=\mathrm{DSPACE}(2^{\mathrm{polynomial}})$ are of particular interest.

We use the following classes of functions.

$$\mathrm{all}=\{f\mid f: \Sigma^*\to\Sigma^*\}$$

$$\mathrm{rec}=\{f\in\mathrm{all}\mid f \text{ is recursive}\}$$

$$\mathrm{p}_i=\{f\in\mathrm{all}\mid f \text{ is computable in } G_i \text{ time}\} \qquad (i\geqslant 1)$$

$$\mathrm{p}_i\mathrm{space}=\{f\in\mathrm{all}\mid f \text{ is computable in } G_i \text{ space}\} \qquad (i\geqslant 1).$$

(The length of the output *is* included as part of the space used in computing f.) We

write p for p_1 and pspace for $p_1$space. Throughout this paper, Δ and Δ' denote one of the classes all, rec, $p_i (i \geq 1)$, $p_i$space$(i \geq 1)$.

A *constructor* is a function $\delta: \{0,1\}^* \to \{0,1\}^*$ that satisfies $x \sqsubsetneq \delta(x)$ for all x. The *result* of a constructor δ (i.e., the language *constructed* by δ) is the unique language $R(\delta)$ such that $\delta^n(\lambda) \sqsubseteq R(\delta)$ for all $n \in \mathbf{N}$. Intuitively, δ constructs $R(\delta)$ by starting with λ and then iteratively generating successively longer prefixes of $R(\delta)$. We write $R(\Delta)$ for the set of languages $R(\delta)$ such that δ is a constructor in Δ. The following routine lemma is the reason for our interest in the above-defined classes of functions.

LEMMA 2.3 [24].

(1) $R(\text{all}) = \mathscr{P}(\{0,1\}^*) = \{0,1\}^\infty$.

(2) $R(\text{rec}) = \text{REC}$.

(3) For $i \geq 1$, $R(p_i) = E_i$.

(4) For $i \geq 1$, $R(p_i\text{space}) = E_i\text{SPACE}$.

Some of our results involve the convergence/divergence of infinite series. A series $\sum_{n=0}^{\infty} a_n$ of nonnegative real numbers a_n is *Δ-convergent* if there is a function $m: \mathbf{N} \to \mathbf{N}$ such that $m \in \Delta$ and

$$\sum_{n=m(i)}^{\infty} a_n \leq 2^{-i}$$

for all $i \in \mathbf{N}$. Such a function m is sometimes called a *modulus* of the convergence. If $\Delta = \text{all}$, this is the usual notion of convergence. If Δ is a time- or space-bounded class of transductions, then Δ-convergence is a stronger condition than convergence. Note that a series is p_i-convergent if and only if it is $p_i$space-convergent.

Adding a layer of uniformity, a sequence

$$\sum_{k=0}^{\infty} a_{j,k} \qquad (j = 0, 1, 2, \ldots)$$

of series of nonnegative real numbers is *uniformly Δ-convergent* if there is a function $m: \mathbf{N}^2 \to \mathbf{N}$ such that $m \in \Delta$ and, for all $j \in \mathbf{N}$, m_j is a modulus of the convergence of the series $\sum_{k=0}^{\infty} a_{j,k}$.

3. RESOURCE-BOUNDED MEASURE

In this section we present those aspects of resource-bounded measure that will be used in the sequel. The formulation here, based on uniform systems of density functions, is much more powerful than the "modulated covering" formulation of [24].

DEFINITION 3.1. A *density function* is a function $d: \{0,1\}^* \to [0, \infty)$ satisfying

$$d(w) \geqslant \frac{d(w0) + d(w1)}{2} \tag{3.1}$$

for all $w \in \{0,1\}^*$. The *global value* of a density function d is $d(\lambda)$. The *set covered by* a density function d is

$$S[d] = \bigcup_{\substack{w \in \{0,1\}^* \\ d(w) \geqslant 1}} C_w. \tag{3.2}$$

A density function d *covers* a set $X \subseteq \{0,1\}^\infty$ if $X \subseteq S[d]$.

For all density functions in this paper, equality actually holds in (3.1) above, but this is not required.

We frequently use the easily verified fact that

$$d(w) \leqslant 2^{|w|}\, d(\lambda) \tag{3.3}$$

holds for all $w \in \{0,1\}^*$ whenever d is a density function.

Consider the random experiment in which a sequence $x \in \{0,1\}^\infty$ is chosen by using an independent toss of a fair coin to decide each bit of x. Taken together, (3.1) and (3.2) imply that $\Pr[x \in S[d]] \leqslant d(\lambda)$ in this experiment. Intuitively, we regard a density function d as a "detailed verification" that $\Pr[x \in X] \leqslant d(\lambda)$ for all sets $X \subseteq S[d]$.

More generally, we are interested in "uniform systems" of density functions that are computable within some resource bound Δ.

DEFINITION 3.2. An n-dimensional *density system* (n-*DS*) is a function

$$d: \mathbf{N}^n \times \{0,1\}^* \to [0, \infty)$$

such that $d_{\vec{k}}$ is a density function for every $\vec{k} \in \mathbf{N}^n$. It is sometimes convenient to regard a density function as a 0-DS.

DEFINITION 3.3. A *computation* of an n-DS d is a function $\hat{d}: \mathbf{N}^{n+1} \times \{0,1\}^* \to \mathbf{D}$ such that

$$|\hat{d}_{\vec{k},r}(w) - d_{\vec{k}}(w)| \leqslant 2^{-r}$$

for all $\vec{k} \in \mathbf{N}^n$, $r \in \mathbf{N}$, and $w \in \{0,1\}^*$. A *Δ-computation* of an n-DS d is a computation $\hat{d}$ of d such that $\hat{d} \in \Delta$. An n-DS d is *Δ-computable* if there exists a Δ-computation $\hat{d}$ of d.

If d is an n-DS such that $d: \mathbf{N}^n \times \{0,1\}^* \to \mathbf{D}$ and $d \in \Delta$, then d is trivially Δ-computable. This fortunate circumstance, in which there is no need to compute approximations, occurs frequently in practice. In any case, we sometimes abuse

notation by writing d for $\hat{d}$, relying on context and subscripts to distinguish an n-DS d from a computation d of d.

We now come to the key idea of resource-bounded measure theory.

DEFINITION 3.4. A *null cover* of a set $X \subseteq \{0,1\}^\infty$ is a 1-DS d such that, for all $k \in \mathbf{N}$, d_k covers X with global value $d_k(\lambda) \leq 2^{-k}$. A *Δ-null cover* of X is a null cover of X that is Δ-computable.

In other words, a null cover of X is a uniform system of density functions that cover X with rapidly vanishing global value. It is easy to show that a set $X \subseteq \{0,1\}^\infty$ has classical Lebesgue measure 0 (i.e., probability 0 in the above coin-tossing experiment) if and only if there exists a null cover of X.

DEFINITION 3.5. A set X has *Δ-measure 0*, and we write $\mu_\Delta(X) = 0$, if there exists a Δ-null cover of X. A set X has *Δ-measure 1*, and we write $\mu_\Delta(X) = 1$, if $\mu_\Delta(X^c) = 0$.

Thus a set X has Δ-measure 0 if Δ provides sufficient computational resources to compute uniformly good approximations to a system of density functions that cover X with rapidly vanishing global value.

We illustrate Definitions 3.4 and 3.5 with a trivial example. (More interesting applications come later, when more machinery is available.)

EXAMPLE 3.6. Let

$$\mathrm{ODD} = \{A \subseteq \{0,1\}^* \mid (\forall n \in \mathbf{N})\, |A_{=n}| \text{ is odd}\}.$$

Define $d: \mathbf{N} \times \{0,1\}^* \to \mathbf{D}$ by the following recursion: For all $k \in \mathbf{N}$, $d_k(\lambda) = 2^{-k}$. If $w \in \{0,1\}^*$ and $|w|$ is *not* of the form $2^{n+1} - 2$ for some $n \in \mathbf{N}$, then $d_k(w0) = d_k(w1) = d_k(w)$. If $w \in \{0,1\}^*$, $b \in \{0,1\}$, and $|w| = 2^{n+1} - 2$, where $n \in \mathbf{N}$, then

$$d_k(wb) = \begin{cases} 2d_k(w) & \text{if } \#(1, w[2^n - 1, 2^{n+1} - 3]) \equiv b \bmod 2 \\ 0 & \text{otherwise.} \end{cases}$$

It is a routine exercise to check that d is a p-null cover of ODD, whence $\mu_{\mathrm{p}}(\mathrm{ODD}) = 0$.

As we have already noted, if $\Delta = \mathrm{all}$, then the Δ-measure 0 sets are precisely the sets of classical Lebesgue measure 0. (Accordingly, we usually write $\mu(X)$ instead of $\mu_{\mathrm{all}}(X)$.) Here we are primarily interested in the internal measure-theoretic structure of complexity classes $R(\Delta)$.

DEFINITION 3.7. A set X has *measure 0 in $R(\Delta)$*, and we write $\mu(X \mid R(\Delta)) = 0$, if $\mu_\Delta(X \cap R(\Delta)) = 0$. A set X has *measure 1 in $R(\Delta)$*, and we write $\mu(X \mid R(\Delta)) = 1$, if $\mu(X^c \mid R(\Delta)) = 0$. If $\mu(X \mid R(\Delta)) = 1$, we say that *almost every* language in $R(\Delta)$ is in X.

If $\Delta = \text{all}$, then $R(\Delta) = \{0,1\}^\infty$, so the conditions $\mu_{\text{all}}(X) = 0$ and $\mu(X \mid R(\Delta)) = 0$ are equivalent to each other and, as we have seen, to the classical condition $\mu(X) = 0$.

If $\Delta = \text{rec}$, then the sets of measure 0 in $R(\Delta) = \text{rec}$ given by Definition 3.7 include all the *effective* measure 0 subsets of REC investigated by Freidzon [7], Mehlhorn [31], and others.

The following lemma is obvious but useful.

LEMMA 3.8. *Let* $X \subseteq \{0,1\}^\infty$.

(a) *If* $\mu_\Delta(X) = 0$ *and* $\Delta \subseteq \Delta'$, *then* $\mu_{\Delta'}(X) = 0$.

(b) *If* $\mu_\Delta(X) = 0$, *then* $\mu(X \mid R(\Delta)) = 0$.

Lemma 3.8 unifies results for various Δ. For example, it gives us the following implications for every set X.

$$\begin{array}{ccccccc} \mu_{\text{p}}(X)=0 & \Longrightarrow & \mu_{\text{pspace}}(X)=0 & \Longrightarrow & \mu_{\text{rec}}(X)=0 & \Longrightarrow & \mu(X)=0 \\ \Downarrow & & \Downarrow & & \Downarrow & & \\ \mu(X \mid \text{E})=0 & & \mu(X \mid \text{ESPACE})=0 & & \mu(X \mid \text{REC})=0 & & \end{array}$$

Thus a proof that a set X has p-measure 0 gives information about the size of X in E and also in larger classes. For example, we saw in Example 3.6 that $\mu_{\text{p}}(\text{ODD}) = 0$; it follows immediately by Lemma 3.8(b) that $\mu(\text{ODD} \mid \text{E}) = 0$. We will see that this means that ODD is a very small subset of E, i.e., that "typical" sequences in E are not elements of ODD. By Lemma 3.8, this also holds if E is replaced by E_2, ESPACE, REC, or $\{0,1\}^\infty$.

In general, if a set X has measure 0 in a class $R(\Delta)$, we interpret this to mean that $X \cap R(\Delta)$ is a "small" subset of $R(\Delta)$. Stated intuitively and simplistically, this interpretation has the following three components.

(s1) Measure 0 sets behave set-theoretically as small sets.

(s2) Very small sets have measure 0.

(s3) Large sets do not have measure 0.

We now develop these points in turn.

For point (s1) we need the following computational restriction of the notion of "countable union."

DEFINITION 3.9. Let $X, X_0, X_1, X_2, \ldots \subseteq \{0,1\}^\infty$.

(a) X is a *Δ-union of the Δ-measure 0 sets* $X_0, X_1, X_2, \ldots$ if $X = \bigcup_{j=0}^\infty X_j$ and there exists a Δ-computable 2-DS d such that each d_j is a null cover of X_j.

(b) X is a *Δ-union of the sets* $X_0, X_1, X_2, \ldots$ *of measure 0 in* $R(\Delta)$ if $X = \bigcup_{j=0}^\infty X_j$ and there exists a Δ-computable 2-DS d such that each d_j is a null cover of $X_j \cap R(\Delta)$.

We now show that the Δ-measure 0 sets and the sets of measure 0 in $R(\Delta)$ are closed under subsets, finite unions, and Δ-unions.

LEMMA 3.10 (Δ-Ideal Lemma). *Let $\mathscr{I}$ be either the collection $\mathscr{I}_\Delta$ of all Δ-measure 0 sets or the collection $\mathscr{I}_{R(\Delta)}$ of all sets that have measure 0 in $R(\Delta)$. In either case, $\mathscr{I}$ has the following three closure properties.*

(a) *If $X \subseteq Y \in \mathscr{I}$, then $X \in \mathscr{I}$.*

(b) *If X is a finite union of elements of $\mathscr{I}$, then $X \in \mathscr{I}$.*

(c) *If X is a Δ-union of elements of $\mathscr{I}$, then $X \in \mathscr{I}$.*

Proof. Property (a) is obvious. It is also obvious that property (b) follows from property (c), since every finite union of elements of $\mathscr{I}$ is trivially a Δ-union of elements of $\mathscr{I}$. It thus suffices to prove (c). In fact, it suffices to prove (c) in the case $\mathscr{I} = \mathscr{I}_\Delta$, since it is easy to see that the case $\mathscr{I} = \mathscr{I}_{R(\Delta)}$ follows directly from this.

So assume that X is a Δ-union of the Δ-measure 0 sets $X_0, X_1, X_2, \ldots$. Then $X = \bigcup_{j=0}^{\infty} X_j$ and there is a Δ-computable 2-DS d such that each $d_{j,k}$ covers X_j with global value $d_{j,k}(\lambda) \leqslant 2^{-k}$. Define a function $d' : \mathbf{N} \times \{0,1\}^* \to [0, \infty)$ by

$$d'_k(w) = \sum_{j=0}^{\infty} d_{j,k+j+1}(w).$$

Each d'_k is, trivially by linearity, a density function, so d' is a 1-DS. We show that d' is a Δ-null cover of X.

To see that each d'_k covers X, fix $k \in \mathbf{N}$ and let $x \in X$. Since $X = \bigcup_{j=0}^{\infty} X_j$ and each $d_{j,k+j+1}$ covers X_j, there exist $j_0, n_0 \in \mathbf{N}$ such that $x \in X_{j_0} \subseteq S[d_{j_0,k+j_0+1}]$ and $d_{j_0,k+j_0+1}(x[0..n_0-1]) \geqslant 1$. We then have

$$\begin{aligned} d'_k(x[0..n_0-1]) &= \sum_{j=0}^{\infty} d_{j,k+j+1}(x[0..n_0-1]) \\ &\geqslant d_{j_0,k+j_0+1}(x[0..n_0-1]) \geqslant 1, \end{aligned}$$

so $x \in S[d'_k]$. Since each d'_k has global value

$$d'_k(\lambda) = \sum_{j=0}^{\infty} d_{j,k+j+1}(\lambda) \leqslant \sum_{j=0}^{\infty} 2^{-(k+j+1)} = 2^{-k},$$

it follows that d' is a null cover of X.

All that remains to be shown is that d' is Δ-computable. For this, let d be a Δ-computation of the 2-DS d. Define the function $d' : \mathbf{N}^2 \times \{0,1\}^* \to \mathbf{D}$ by

$$d'_{k,r}(w) = \sum_{j=0}^{r+|w|} d_{j,k+j+1,r+j+2}(w).$$

We show that d' is a Δ-computation of the 2-DS d'. It is clear that $d' \in \Delta$. Letting $\sigma = \sum_{j=0}^{r+|w|} d_{j,k+j+1}(w)$, we have

$$|d'_{k,r}(w) - \sigma| \leqslant \sum_{j=0}^{r+|w|} |d_{j,k+j+1,r+j+2}(w) - d_{j,k+j+1}(w)|$$
$$\leqslant \sum_{j=0}^{r+|w|} 2^{-(r+j+2)} \leqslant \sum_{j=0}^{\infty} 2^{-(r+j+2)} = 2^{-(r+1)}$$

and, by (3.3),

$$|\sigma - d'_k(w)| = \sum_{j=r+|w|+1}^{\infty} d_{j,k+j+1}(w)$$
$$\leqslant \sum_{j=r+|w|+1}^{\infty} 2^{|w|}\, d_{j,k+j+1}(\lambda)$$
$$\leqslant 2^{|w|} \sum_{j=r+|w|+1}^{\infty} 2^{-(k+j+1)}$$
$$= 2^{-(k+r+1)} \leqslant 2^{-(r+1)},$$

so

$$|d'_{k,r}(w) - d'_k(w)| \leqslant 2^{-r}$$

for all $k, r \in \mathbf{N}$ and $w \in \{0,1\}^*$. Thus d' is a Δ-computation of d' and the proof is complete. ∎

In the classical case, where $\Delta = \text{all}$, a Δ-union is simply a countable union and Lemma 3.10 tells us that the measure 0 sets are closed under subsets, finite unions, and countable unions. This well-known fact is usually expressed by saying that the measure 0 sets form a *σ-ideal* of subsets of $\{0,1\}^\infty$. Extending this terminology, we conclude from Lemma 3.10 that the Δ-measure 0 sets form a *Δ-ideal* of subsets of $\{0,1\}^\infty$ and that the measure 0 subsets of $R(\Delta)$ form a *Δ-ideal* of subsets of $R(\Delta)$. This is the precise formulation of point (s1).

For point (s2) we define a computationally restricted notion of "countable set."

DEFINITION 3.11. A set $X \subseteq R(\Delta)$ is *Δ-countable* if there is a function $\delta: \mathbf{N} \times \{0,1\}^* \to \{0,1\}^*$ such that $\delta \in \Delta$, δ_k is a constructor for each $k \in \mathbf{N}$, and $X = \{R(\delta_k) \mid k \in \mathbf{N}\}$.

LEMMA 3.12. *Let* $X \subseteq R(\Delta)$.

(a) *If X is finite, then* $\mu_\Delta(X) = 0$.

(b) *If X is Δ-countable, then* $\mu_\Delta(X) = 0$.

Proof. Since finite subsets of $R(\Delta)$ are trivially Δ-countable, it suffices to prove (b). So let $\delta \in \Delta$ testify that $X \subseteq R(\Delta)$ is Δ-countable. Define $d: \mathbf{N}^2 \times \{0,1\}^* \to \mathbf{D}$ by

$$d_{k,l}(w) = 2^{m-l},$$

where $m \in \mathbf{N}$ is greatest such that $\delta_k^m(\lambda) \sqsubseteq w$. It is clear that $d \in \Delta$ and that each d_k is a null cover of the singleton set $\{R(\delta_k)\}$. That is, d testifies that X is a Δ-union of the Δ-measure 0 sets $\{R(\delta_k)\}$. It follows by the Δ-Ideal Lemma that $\mu_\Delta(X) = 0$. ∎

Lemma 3.12 is our precise formulation of point (s2). In particular it implies that every singleton subset $\{x\}$ of $R(\Delta)$ has Δ-measure 0 (hence measure 0 in $R(\Delta)$). It should be noted that the assumption that $\{x\} \subseteq R(\Delta)$ cannot be deleted here. We see in Section 6 that arbitrary singleton sets $\{x\}$ may fail to have Δ-measure 0.

We now come to point (s3). This is the most crucial issue in our development. If we are to endow a complexity class $R(\Delta)$ with internal measure-theoretic structure, then $R(\Delta)$ itself must be a large set, hence by (s3) must not have measure 0 in $R(\Delta)$. That is, the Δ-ideal $\mathscr{I}_{R(\Delta)}$ of all measure 0 subsets of $R(\Delta)$ must be *proper* in the sense that $R(\Delta) \notin \mathscr{I}_{R(\Delta)}$. In cases of interest, $R(\Delta)$ is a countable set and thus has classical measure 0. Fortunately, however, $R(\Delta)$ does not have Δ-measure 0. This fact follows from the following conservation principle, which says that, within the computational resources of Δ, the intersection of a cylinder with $R(\Delta)$ cannot be covered more economically than the cylinder itself. (Recall that, for $z \in \Sigma^*$, $\mu(z) = 2^{-|z|}$ is the measure of the cylinder C_z.)

THEOREM 3.13 (Measure Conservation Theorem). *If C_z is a cylinder and d is a Δ-computable density function that covers $C_z \cap R(\Delta)$, then $d(\lambda) \geq \mu(z)$.*

Proof. Assume that d is a Δ-computable density function such that $d(\lambda) < \mu(z)$. We prove by diagonalization that d does not cover $C_z \cap R(\Delta)$. Specifically, we exhibit a constructor $\delta \in \Delta$ such that

$$z \sqsubseteq R(\delta), \tag{3.4}$$

$$|\delta(x)| = |x| + 1 \qquad \text{for all} \quad x \in \{0,1\}^*, \tag{3.5}$$

and

$$d(\delta^k(\lambda)) < 1 \qquad \text{for all} \quad k \in \mathbf{N}. \tag{3.6}$$

(It follows from these three things that

$$R(\delta) \in C_z \backslash S[d],$$

whence d does not cover $C_z \cap R(\Delta)$.)

Let $m = \max\{1, |z|\}$ and let

$$S = \{y \in \{0,1\}^m \mid z \sqsubseteq y\}.$$

(We emphasize that $z \in \{0,1,\perp\}^*$, $\|z\| \leq |z| \leq m$, and $S \subseteq \{0,1\}^m$.) For each $y \in S$, let $g(y) \in \{0,1\}^{\leq m}$ be the shortest prefix of y such that, for every prefix w of y, $d(w) \leq d(g(y))$. We first note that there exists $y \in S$ such that

$$d(g(y)) \leq \frac{d(\lambda)}{\mu(z)}. \tag{3.7}$$

To see this, define d': $\{0,1\}^* \to [0,\infty)$ by

$$d'(x) = \begin{cases} d(g(y)) & \text{if } y \in S \text{ and } g(y) \sqsubseteq x \\ d(x) & \text{if no element of } g(S) \text{ is a prefix of } x. \end{cases}$$

(The function d' is well-defined because $g(S)$ is an instantaneous code; i.e., no element of $g(S)$ is a prefix of any other.) It is readily checked that d' is a density function, so

$$\begin{aligned} d(\lambda) = d'(\lambda) &\geqslant 2^{-m} \sum_{y \in \{0,1\}^m} d'(y) \geqslant 2^{-m} \sum_{y \in S} d'(y) \\ &\geqslant 2^{-m} |S| \min_{y \in S} d'(y) = 2^{-|z|} \min_{y \in S} d'(y) \\ &= \mu(z) \min_{y \in S} d(g(y)), \end{aligned}$$

so some $y \in S$ satisfies (3.7).

Fix $q \in \mathbf{D}$ and a positive integer l such that

$$d(\lambda) \leqslant q \cdot \mu(z), \qquad q + 2^{1-l} \leqslant 1. \tag{3.8}$$

Let d be a Δ-computation of the density function (i.e., 0-DS) d. Using d and the constants m, y, q, and l, define the constructor δ: $\{0,1\}^* \to \{0,1\}^*$ by

$$\delta(x) = \begin{cases} xy[|x|] & \text{if } x \sqsubsetneq y \\ x0 & \text{if } d_{a(x)}(x0) \leqslant d_{a(x)}(x) + 2^{1-a(x)} \text{ and not } x \sqsubsetneq y \\ x1 & \text{otherwise,} \end{cases}$$

where $a(x) = |x| + l + 3$. It is clear that $\delta \in \Delta$ and that (3.5) holds. Also, $z \sqsubseteq y = \delta^m(\lambda)$, so (3.4) holds. All that remains, then, is to verify (3.6).

A key property of δ is that

$$d_{a(x)}(\delta(x)) \leqslant d_{a(x)}(x) + 2^{1-a(x)} \tag{3.9}$$

holds whenever x is not a proper prefix of y. To see this, we need only recall that d is a density function, whence $d_{a(x)}(x0) > d_{a(x)}(x) + 2^{1-a(x)}$ implies that

$$\begin{aligned} d_{a(x)}(x1) &\leqslant d(x1) + 2^{-a(x)} \leqslant 2d(x) - d(x0) + 2^{-a(x)} \\ &\leqslant 2d_{a(x)}(x) - d_{a(x)}(x0) + 2^{2-a(x)} < d_{a(x)}(x) + 2^{1-a(x)}. \end{aligned}$$

By (3.7) and (3.8) we have, for all $0 \leqslant k \leqslant m$,

$$\begin{aligned} d(\delta^k(\lambda)) = d(y[0..k-1]) \leqslant d(g(y)) &\leqslant \frac{d(\lambda)}{\mu(z)} \leqslant q \\ &< q + 2^{-l}(1 - 2^{-|\delta^k(\lambda)|}). \end{aligned} \tag{3.10}$$

Also, if $x \in \{0,1\}^*$ is such that $d(x) < q + 2^{-l}(1 - 2^{-|x|})$ and x is not a proper prefix of y, then (3.9) ensures that

$$\begin{aligned} d(\delta(x)) &\leq d_{a(x)}(\delta(x)) + 2^{-a(x)} \leq d_{a(x)}(x) + 3 \cdot 2^{-a(x)} \leq d(x) + 2^{2-a(x)} \\ &< q + 2^{-l}(1 - 2^{-|x|}) + 2^{-(|x|+l+1)} = q + 2^{-l}(1 - 2^{-|\delta(x)|}). \end{aligned} \tag{3.11}$$

Taken together, (3.10) and (3.11) provide an inductive proof that

$$d(\delta^k(\lambda)) < q + 2^{-l}(1 - 2^{-|\delta^k(\lambda)|}) < q + 2^{-l} \leq 1$$

for all $k \in \mathbf{N}$; i.e., (3.6) holds. This completes the proof. ∎

COROLLARY 3.14. *The Δ-ideals $\mathscr{I}_\Delta$ and $\mathscr{I}_{R(\Delta)}$ of Lemma* 3.10 *are both proper. In fact, neither of these Δ-ideals contains $C_z \cap R(\Delta)$ for any nonempty cylinder C_z.*

The implications

$$\mathscr{I}_{R(\Delta)} \text{ is proper} \Rightarrow \mathscr{I}_\Delta \text{ is proper}$$

and

$$\mathscr{I}_{\text{all}} \text{ is proper} \Rightarrow \mathscr{I}_\Delta \text{ is proper}$$

are both trivial, and Borel proved long ago (using a classical version of Theorem 3.13) that $\mathscr{I}_{\text{all}}$ is proper, i.e., that not every set has measure 0. The real content of Corollary 3.14 is the assertion that $\mathscr{I}_{R(\Delta)}$ is proper, i.e., that (s3) holds internally for the classes $R(\Delta)$.

This completes the interpretation of measure 0 sets as small sets. We now give a useful criterion for proving that sets have Δ-measure 0. This theorem is a uniform, resource-bounded extension of the classical first Borel–Cantelli lemma.

THEOREM 3.15. *If d is a Δ-computable 2-DS such that the series*

$$\sum_{k=0}^{\infty} d_{j,k}(\lambda) \qquad (j = 0, 1, 2, \ldots) \tag{3.12}$$

are uniformly Δ-convergent, then

$$\mu_\Delta\left(\bigcup_{j=0}^{\infty} \bigcap_{t=0}^{\infty} \bigcup_{k=t}^{\infty} S[d_{j,k}]\right) = 0.$$

The coordinate j of Theorem 3.15 is often not needed in applications. Discarding this layer of uniformity gives the following simplification.

COROLLARY 3.16. *If d is a Δ-computable 1-DS such that the series*

$$\sum_{k=0}^{\infty} d_k(\lambda)$$

is Δ-convergent, then

$$\mu_{\Delta}\left(\bigcap_{t=0}^{\infty}\bigcup_{k=t}^{\infty} S[d_k]\right)=\mu_{\Delta}(\{x\in\{0,1\}^{\infty}\mid x\in S[d_k]\ i.o.\})=0.$$

Before proving Theorem 3.15 we give a simple example of its use.

EXAMPLE 3.17. Fix a real number $0<\varepsilon<1$ and let

$$X=\{x\in\{0,1\}^{\infty}\mid x[k..k+\lfloor k^{\varepsilon}\rfloor]\in\{0\}^{*}\ \text{i.o.}\}.$$

Let $l=\lceil 1/\varepsilon\rceil$ and define $d\colon \mathbf{N}\times\{0,1\}^{*}\to\mathbf{D}$ by the following recursion. If $|w|\leqslant k$, then $d_k(w)=2^{-\lfloor k^{1/l}\rfloor}$. If $|w|\geqslant k$, then $d_k(w0)=2d_k(w)$ and $d_k(w1)=0$. Then, for all $x\in\{0,1\}^{\infty}$ and $k\in\mathbf{N}$,

$$x[k..k+\lfloor k^{\varepsilon}\rfloor]\in\{0\}^{*}\Rightarrow d_k(x[0..k+\lfloor k^{\varepsilon}\rfloor])=2^{\lfloor k^{\varepsilon}\rfloor-\lfloor k^{1/l}\rfloor}\geqslant 1$$

$$\Rightarrow x\in S[d_k],$$

so $X\subseteq\bigcap_{t=0}^{\infty}\bigcup_{k=t}^{\infty}S[d_k]$. Since $d\in\mathrm{p}$ and the series $\sum_{k=0}^{\infty}d_k(\lambda)=\sum_{k=0}^{\infty}2^{-\lfloor k^{1/l}\rfloor}$ is, by routine calculus, p-convergent, it follows by Corollary 3.16 that $\mu_{\mathrm{p}}(X)=\mu(X\mid\mathrm{E})=0$. That is, for every $\varepsilon>0$, for almost every sequence $x\in\mathrm{E}$, there are at most finitely many k for which $x[k..k+\lfloor k^{\varepsilon}\rfloor]$ consists entirely of zeroes.

Proof of Theorem 3.15. Assume the hypothesis. Fix a function $m\colon\mathbf{N}^2\to\mathbf{N}$ testifying that the series (3.12) are uniformly Δ-convergent. Without loss of generality, assume that m_j is nondecreasing and $m_j(n)\geqslant 2$ for all $j,n\in\mathbf{N}$. Define

$$S_{j,t}=\bigcup_{k=t}^{\infty}S[d_{j,k}],$$

$$S_j=\bigcap_{t=0}^{\infty}S_{j,t},$$

$$S=\bigcup_{j=0}^{\infty}S_j.$$

Our task is to prove that $\mu_{\Delta}(S)=0$. Define $d'\colon\mathbf{N}^2\times\{0,1\}^{*}\to[0,\infty)$ by

$$d'_{j,n}(w)=\sum_{k=m_j(n)}^{\infty}d_{j,k}(w)$$

for all $j,n\in\mathbf{N}$ and $w\in\{0,1\}^{*}$. We show that d' testifies that S is a Δ-union of the Δ-measure 0 sets $S_0,S_1,S_2,\ldots$, whence $\mu_{\Delta}(S)=0$ by the Δ-Ideal Lemma.

Each $d'_{j,n}$ is trivially by linearity a density function, so d' is a 2-DS. To see that each d'_j is a null cover of S_j, fix $j,n\in\mathbf{N}$. Let $x\in S_j$. Then $x\in\bigcap_{t=0}^{\infty}S_{j,t}$, so

$$x\in S_{j,m_j(n)}=\bigcup_{k=m_j(n)}^{\infty}S[d_{j,k}],$$

so there exist $k_0 \geq m_j(n)$ and $l_0 \in \mathbf{N}$ such that $d_{j,k_0}(x[0..l_0-1]) \geq 1$. We then have

$$\begin{aligned} d'_{j,n}(x[0..l_0-1]) &= \sum_{k=m_j(n)}^{\infty} d_{j,k}(x[0..l_0-1]) \\ &\geq d_{j,k_0}(x[0..l_0-1]) \geq 1, \end{aligned}$$

so $x \in S[d'_{j,n}]$. Thus $d'_{j,n}$ covers S_j. Moreover, the global value of $d'_{j,n}$ satisfies

$$d'_{j,n}(\lambda) = \sum_{k=m_j(n)}^{\infty} d_{j,k}(\lambda) \leq 2^{-n}.$$

Thus each d'_j is a null cover of S_j.

It remains to be shown that the 2-DS d' is Δ-computable. For this, let d be a Δ-computation of the 2-DS d. Define $d'\colon \mathbf{N}^3 \times \{0,1\}^* \to \mathbf{D}$ by

$$d'_{j,n,r}(w) = \sum_{k=m_j(n)}^{m_j(r+|w|+1)} d_{j,k,r+k}(w).$$

It is clear that $d' \in \Delta$. Fix $j, n, r \in \mathbf{N}$ and $x \in \{0,1\}^*$. Let $\sigma = \sum_{k=m_j(n)}^{m_j(r+|w|+1)} d_{j,k}(w)$. Then

$$\begin{aligned} |d'_{j,n,r}(w) - \sigma| &\leq \sum_{k=m_j(n)}^{m_j(r+|w|+1)} |d_{j,k,r+k}(w) - d_{j,k}(w)| \\ &\leq \sum_{k=m_j(n)}^{m_j(r+|w|+1)} 2^{-(r+k)} \\ &\leq \sum_{k=m_j(n)}^{\infty} 2^{-(r+k)} \\ &= 2^{1-(r+m_j(n))} \leq 2^{-(r+1)} \end{aligned}$$

and, by (3.3),

$$\begin{aligned} |d'_{j,n}(w) - \sigma| &\leq \sum_{k=m_j(r+|w|+1)}^{\infty} d_{j,k}(w) \\ &\leq 2^{|w|} \sum_{k=m_j(r+|w|+1)}^{\infty} d_{j,k}(\lambda) \\ &\leq 2^{-(r+1)}, \end{aligned}$$

so

$$|d'_{j,n,r}(w) - d'_{j,n}(w)| \leq 2^{-r}.$$

Thus d' is a Δ-computation of the 2-DS d'. ∎

Individually, the density functions used here closely resemble the martingales used by Schnorr [34–37] in his investigation of random and pseudorandom sequences. Indeed, a *martingale*, as defined by Schnorr, is formally a density function satisfying (3.1) with equality. This equality requirement does not make any difference to his work or ours, so density functions and martingales have essentially identical formal definitions. There is, however, substantial difference in the spirit and use of these two notions. Schnorr, following early work of Ville, used martingales to formalize the notion of variable-stakes gambling strategies. In this context, one is typically interested in ideas of the following sort.

DEFINITION 3.18. A martingale d *succeeds* on a sequence $x \in \{0,1\}^\infty$ if

$$\limsup_{n\to\infty} d(x[0..n-1]) = \infty.$$

Schnorr, using technical variants of Definition 3.18 (strong success notions involving the rate of growth of the lim sup), has shown that the "weak failure" of all *individual* Δ-computable martingales on a sequence x characterizes a weak pseudorandomness condition [34, 36]. (See also [41, 42] and Section 6 below.)

In contrast, the density functions here are generalizations of the density function d of [24, Lemma 5.8]. We have first used *uniform systems* of such density functions to define resource-bounded measure and only then used resource-bounded measure to define pseudorandomness. (See Section 6 below and [27].) This is a natural development in investigating the internal, measure-theoretic structure of complexity classes.

In [26], the *Δ-measurability* of sets $X \subseteq \{0,1\}^\infty$ and the *measure* $\mu_\Delta(X)$ of Δ-measurable sets ($0 \leqslant \mu_\Delta(X) \leqslant 1$) are defined and developed in terms of uniform systems of density functions. Definition 3.5 above is a special case (the measure zero/one case) of these definitions. As it turns out, individual martingales can be used to characterize this special case:

THEOREM 3.19. *A set $X \subseteq \{0,1\}^\infty$ has Δ-measure 0 if and only if there exists a Δ-computable martingale d that succeeds on every sequence $x \in X$.*

(We do not use Theorem 3.19 in this paper. The proof will appear in [26].)

Notwithstanding the contrast between our approach and his, we emphasize that many technical aspects of Section 3 (e.g., much of the content of the Measure Conservation Theorem) were already present, some 20 years ago, in the work of C. P. Schnorr.

4. KOLMOGOROV COMPLEXITY

In this, the main section of the paper, we prove that almost every initial segment of almost every binary sequence computable in exponential resources has very high

resource-bounded Kolmogorov complexity. Of course we must first formulate this assertion more precisely.

In order to make our lower bounds applicable to other complexity criteria (e.g., the circuit-size lower bounds in Section 5), we introduce a new generalization of Kolmogorov complexity, called *selective* Kolmogorov complexity. We then focus on the space- and time-bounded selective Kolmogorov complexities of infinite binary sequences.

Some terminology and notation are useful. For a fixed machine M and "program" $\pi \in \{0,1\}^*$ for M, if $M(\langle \pi, 0^n \rangle)$ halts with output $w \in \{0,1\}^n$, then we write $M(\pi, n)$ for the binary string w. In particular, an assertion that $M(\pi, n)$ has some particular property "in $\leqslant t$ time" (respectively, "in $\leqslant t$ space") means that $M(\langle \pi, 0^n \rangle)$ halts with an output string $M(\pi, n) \in \{0,1\}^n$ in $\leqslant t$ steps (respectively, using $\leqslant t$ space) and that this output string has the indicated property. Note that this notation implicitly requires $M(\pi, n)$ to be a binary string whose length is exactly n.

DEFINITION 4.1. A *selector* is a function $\sigma: \mathbf{N} \to \{\bot, \top\}^*$ such that $|\sigma(n)| = n$ for each $n \in \mathbf{N}$. We write $\#\sigma(n)$ for the number of occurrences of $\bot$ in $\sigma(n)$.

DEFINITION 4.2. Let M be a machine, let $t: \mathbf{N} \to \mathbf{N}$ be a resource bound, let σ be a selector, and let $x \in \{0,1\}^\infty$.

(a) The *t-time-bounded σ-selective Kolmogorov complexity* of x relative to M is the function $\mathrm{KT}^t_M(x \wedge \sigma | \cdot): \mathbf{N} \to \mathbf{N} \cup \{\infty\}$ defined by

$$\mathrm{KT}^t_M(x \wedge \sigma | n) = \min\{|\pi| \mid M(\pi, n) \sqsubseteq x \wedge \sigma(n) \text{ in } \leqslant t(n) \text{ time}\}.$$

(b) The *t-space-bounded σ-selective Kolmogorov complexity* of x relative to M is the function $\mathrm{KS}^t_M(x \wedge \sigma | \cdot): \mathbf{N} \to \mathbf{N} \cup \{\infty\}$ defined by

$$\mathrm{KS}^t_M(x \wedge \sigma | n) = \min\{|\pi| \mid M(\pi, n) \sqsubseteq x \wedge \sigma(n) \text{ in } \leqslant t(n) \text{ space}\}.$$

Just as for other resource-bounded Kolmogorov complexities (see Huynh [13], for example), well-known simulation techniques show that there exist a universal machine U and a polynomial q such that for each machine M there is a constant c such that for all t, σ, x, and n we have

$$\mathrm{KT}^{q(ct+c)}_U(x \wedge \sigma | n) \leqslant \mathrm{KT}^t_M(x \wedge \sigma | n) + c \tag{4.1}$$

and

$$\mathrm{KS}^{ct+c}_U(x \wedge \sigma | n) \leqslant \mathrm{KS}^t_M(x \wedge \sigma | n) + c. \tag{4.2}$$

As usual, we fix such a universal machine U and omit it from the notation.

The t-time-bounded σ-selective Kolmogorov complexity of a binary sequence x is thus the function $\mathrm{KT}^t(x \wedge \sigma | \cdot)$ whose value at an argument n is the length $\mathrm{KT}^t(x \wedge \sigma | n)$ of the shortest program π such that $U(\pi, n) \sqsubseteq x \wedge \sigma(n)$. The latter condition says that $U(\pi, n)[i]$ must agree with $x[i]$ for every $0 \leqslant i < n$ such that

$\sigma(n)[i] = \perp$. No requirement is placed on $U(\pi, n)[i]$ when $\sigma(n)[i] = \top$. That is (relative to the universal machine U), π must correctly decide x at each of the $\#\sigma(n)$ positions selected by $\sigma(n)$.

If σ is a selector that is computable in polynomial time, then it is easy to design a machine M_σ that, on input $\langle \pi, 0^n \rangle$ with $\pi \in \{0,1\}^{\#\sigma(n)}$, outputs in polynomial time a string $M_\sigma(\pi, n)$ such that, if $i_0 < \cdots < i_{\#\sigma(n)-1}$ are the indices i for which $\sigma(n)[i] = \perp$, then $M_\sigma(\pi, n)[i_0] \cdots M_\sigma(\pi, n)[i_{\#\sigma(n)-1}] = \pi$ and $M_\sigma(\pi, n)[i] = 0$ for all other indices i. Hence $M_\sigma(\pi, n)$ is the n-bit binary string consisting of the program π positioned in $M_\sigma(\pi, n)$ according to σ, with zeroes in all remaining positions. For example, if $\sigma(6) = \top\perp\perp\top\perp\top$ and $\pi = 101$, then $M_\sigma(\pi, 6) = 0\underline{1}00\underline{1}0$, where we have underlined the positions selected by $\sigma(6)$. It is clear that there is a polynomial q such that $\mathrm{KT}^q_{M_\sigma}(x \wedge \sigma \mid n) \leqslant \#\sigma(n)$ for all x and n. It follows by (4.1) that there exist a polynomial q and a constant c such that

$$\mathrm{KT}^q(x \wedge \sigma \mid n) \leqslant \#\sigma(n) + c \tag{4.3}$$

for all $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$. That is, the polynomial time-bounded σ-selective Kolmogorov complexity cannot be much larger than $\#\sigma(n)$, the number of bits to be correctly decided. Note that the polynomial q here depends on the running time of the selector σ but not on x or n.

A similar argument shows that if σ is a selector that is computable in polynomial space, then there exist a polynomial q and a constant c such that

$$\mathrm{KS}^q(x \wedge \sigma \mid n) \leqslant \#\sigma(n) + c \tag{4.4}$$

for all $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$.

As a special case of the selective Kolmogorov complexity, we have the *conditional* Kolmogorov complexity. (This is actually a much-studied special case, adapted to infinite sequences, of the conditional complexity defined by Kolmogorov [18].) Again, we are interested in resource-bounded versions.

DEFINITION 4.3. Let $t: \mathbf{N} \to \mathbf{N}$ be a resource bound and let $x \in \{0,1\}^\infty$.

(a) The *t-time-bounded conditional Kolmogorov complexity* of x is the function $\mathrm{KT}^t(x \mid \cdot) = \mathrm{KT}^t(x \wedge \sigma \mid \cdot)$, where the selector σ is defined by $\sigma(n) = \perp^n$ for all $n \in \mathbf{N}$.

(b) The *t-space-bounded conditional Kolmogorov complexity* of x is the function $\mathrm{KS}^t(x \mid \cdot) = \mathrm{KS}^t(x \wedge \sigma \mid \cdot)$, where σ is as in part (a).

Thus the conditional Kolmogorov complexity is the special case of the selective Kolmogorov complexity in which every position is selected, i.e., every bit of $U(\pi, n)$ must be correct for x.

From (4.3) and (4.4) we get the well-known fact that there exist a polynomial q (which is in fact linear) and a constant c such that

$$\mathrm{KT}^q(x \mid n) \leqslant n + c \tag{4.5}$$

and

$$\mathrm{KS}^q(x \mid n) \leqslant n + c \tag{4.6}$$

hold for all $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$. It is also clear that the inequalities

$$\mathrm{KT}^t(x \wedge \sigma | n) \leqslant \mathrm{KT}^t(x|n) \tag{4.7}$$

and

$$\mathrm{KS}^t(x \wedge \sigma | n) \leqslant \mathrm{KS}^t(x|n) \tag{4.8}$$

hold for all t, σ, x, and n.

Our primary objective in this section is to establish lower bounds that hold almost everywhere in various complexity classes for the time- and space-bounded conditional Kolmogorov complexities. Our secondary objective is to do this in such a manner that the circuit-size lower bounds of Section 5 can then be derived. Accordingly, we prove our lower bounds for the time- and space-bounded selective Kolmogorov complexities. By (4.7) and (4.8), this obviously achieves our primary objective. We see in Section 5 that the secondary objective is also achieved.

We now prove an almost-everywhere lower bound for space-bounded selective program size in ESPACE.

THEOREM 4.4. *Suppose that a selector σ and a function $f: \mathbf{N} \times \mathbf{N}$ have the following properties.*

(i) $\sigma, f \in$ *pspace.*

(ii) $\sum_{n=0}^{\infty} 2^{-f(\#\sigma(n))}$ *is p-convergent.*

Then for every polynomial q, the set of all $x \in \{0,1\}^\infty$ such that

$$KS^q(x \wedge \sigma | n) > \#\sigma(n) - f(\#\sigma(n)) \text{ a.e.}$$

has pspace-measure 1, *hence measure* 1 *in ESPACE.*

Proof. For each $n \in \mathbf{N}$, let

$$X_n = \{x \mid \mathrm{KS}^q(x \wedge \sigma | n) \leqslant \#\sigma(n) - f(\#\sigma(n))\}.$$

It suffices to prove that $\{x \mid x \in X_n \text{ i.o.}\}$ has pspace-measure 0. For this, it suffices by Corollary 3.16 to exhibit a pspace-computable 1-DS d such that each $X_n \subseteq S[d_n]$ and the series $\sum_{n=0}^{\infty} d_n(\lambda)$ is p-convergent.

For each $n \in \mathbf{N}$, let

$$B_n = \{\pi \in \{0,1\}^{\leqslant \#\sigma(n) - f(\#\sigma(n))} \mid U(\pi, n) \in \{0,1\}^n \text{ in } \leqslant q(n) \text{ space}\}$$

and, for all $\pi \in B_n$, let

$$Z_{n,\pi} = \{x \in \{0,1\}^\infty \mid U(\pi, n) \sqsubseteq x \wedge \sigma(n)\}.$$

Define $d: \mathbf{N} \times \{0,1\}^* \to [0, \infty)$ by

$$d_n(w) = \sum_{\pi \in B_n} P(Z_{n,\pi} | C_w), \tag{4.9}$$

where the conditional probability

$$P(Z_{n,\pi} \mid C_w) = \Pr_x [x \in Z_{n,\pi} \mid x \in C_w]$$

is chosen according to the random experiment in which an independent toss of a fair coin is used to decide each bit of a sequence $x \in \{0, 1\}^\infty$.

Since each

$$P(Z_{n,\pi} \mid C_w) = \frac{P(Z_{n,\pi} \mid C_{w0}) + P(Z_{n,\pi} \mid C_{w1})}{2},$$

it is clear that d is a 1-DS. Moreover, for all $n \in \mathbf{N}$, $\pi \in B_n$, and $w \in \{0, 1\}^*$, it is easy to see that

$$P(Z_{n,\pi} \mid C_w) = 2^{|U(\pi,n) \wedge \sigma(n) \wedge w| - |w|}. \tag{4.10}$$

Using (4.9) and (4.10), it is clear that $d \in$ pspace, whence d is certainly pspace-computable.

To see that d has the desired covering property, fix $n \in \mathbf{N}$ and let $x \in X_n$. Then there exists $\pi \in B_n$ such that $x \in Z_{n,\pi}$. For all $y \in C_{x[0..n-1]}$, we then have $U(\pi, n) \sqsubseteq x \wedge \sigma(n) = y \wedge \sigma(n)$, so $C_{x[0..n-1]} \subseteq Z_{n,\pi}$. It follows that

$$d_n(x[0..n-1]) \geqslant P(Z_{n,\pi} \mid C_{x[0..n-1]}) = 1,$$

whence $x \in S[d_n]$. Thus $X_n \subseteq S[d_n]$.

Finally, note that each

$$d_n(\lambda) = \sum_{\pi \in B_n} P(Z_{n,\pi}) = 2^{-\#\sigma(n)} |B_n| < 2^{1 - f(\#\sigma(n))}.$$

Since $\sum_{n=0}^{\infty} 2^{-f(\#\sigma(n))}$ is p-convergent, it follows immediately that $\sum_{n=0}^{\infty} d_n(\lambda)$ is p-convergent. By Corollary 3.16, this completes the proof. ∎

Several results, some new and some previously known, are easily derived from Theorem 4.4 and its proof. We first give almost everywhere lower bounds for space-bounded conditional Kolmogorov complexity.

THEOREM 4.5. *If $f: \mathbf{N} \to \mathbf{N}$, $f \in$ pspace, and the series $\sum_{n=0}^{\infty} 2^{-f(n)}$ is p-convergent, then for every polynomial q, the set of all $x \in \{0, 1\}^\infty$ such that $KS^q(x \mid n) > n - f(n)$ a.e. has pspace-measure 1, hence measure 1 in ESPACE.*

Proof. This follows immediately from Theorem 4.4 if we use the selector σ defined by $\sigma(n) = \perp^n$ for all $n \in \mathbf{N}$. ∎

COROLLARY 4.6. *For every polynomial q and every real number $\varepsilon > 0$, the set of all $x \in \{0, 1\}^\infty$ such that $KS^q(x \mid n) > n - n^\varepsilon$ a.e. has pspace-measure 1, hence measure 1 in ESPACE.*

Proof. Routine calculus shows that the series $\sum_{n=0}^{\infty} 2^{-n^{\epsilon}}$ is p-convergent, so this follows immediately from Theorem 4.5. ▮

Corollary 4.6 immediately implies (in fact, is much stronger than) the following two results, which have been used to investigate complexity properties of problems that are hard for ESPACE under resource-bounded Turing reducibilities.

COROLLARY 4.7 (Huynh [13]). *There is a sequence $x \in ESPACE$ such that $KS^n(x \mid n) > n/4$ a.e.*

COROLLARY 4.8 (Lutz [24]). *For every polynomial q and every real number $\beta < 1$, the set of all $x \in \{0,1\}^{\infty}$ such that $KS^q(x \mid n) > \beta n$ i.o. has pspace-measure 1, hence measure 1 in ESPACE.*

A brief examination of the proof of Theorem 4.4 shows that it remains valid if pspace is replaced by any of the resource bounds Δ for which pspace $\subseteq \Delta$. Moreover, the result continues to hold if the polynomial restriction on q is relaxed, as long as q-space-bounded computation can be carried out within the resources afforded by Δ. Taking $\Delta = \text{rec}$, then, we have the following, which is essentially a weak version of Theorem 4.5.

COROLLARY 4.9. *If $f, g: \mathbf{N} \to \mathbf{N}$ are computable and $\sum_{n=0}^{\infty} 2^{-f(n)}$ is rec-convergent, then the set of all $x \in \{0,1\}^{\infty}$ such that $KS^g(x \mid n) > n - f(n)$ a.e. has rec-measure 1, hence measure 1 in REC.*

Corollary 4.9 says that almost every recursive sequence has very high space-bounded Kolmogorov complexity in almost every initial segment. The following known result follows easily from this.

COROLLARY 4.10 (Ko [17]). *If $f, g: \mathbf{N} \to \mathbf{N}$ are computable and $\sum_{n=0}^{\infty} 2^{-f(n)}$ converges, then there is a recursive sequence $x \in \{0,1\}^{\infty}$ such that $KS^g(x \mid n) > n - f(n) - \log n$ a.e.*

Proof. We just note that if $\sum_{n=0}^{\infty} 2^{-f(n)}$ converges, then $\sum_{n=0}^{\infty} 2^{-f(n)-\log n}$ is rec-convergent. (This is a special case of the following obvious fact. If a series $\sum_{n=0}^{\infty} a_n$ converges and a sequence $\{b_n\}$ Δ-converges to 0, where the a_n and b_n are all non-negative, then the series $\sum_{n=0}^{\infty} a_n b_n$ is Δ-convergent.) The present result thus follows immediately from Corollaries 4.9 and 3.16. ▮

It is worthwhile to pause for a moment and consider the roles played by various methods. Corollaries 4.9 and 4.10 provide a good focal point for this. Our proof of Corollary 4.9 is essentially that of Theorem 4.4, with resource bounds relaxed and selectors removed (i.e., replaced by the selector $\sigma(n) = \perp^n$). With these modifications, the proof is a transparent covering argument, simpler than the Meyer and McCreight [32] weighted priority diagonalization used by Ko [17] to prove Corollary 4.10. Does this give us a new proof of Corollary 4.10, free of the weighted

priority diagonalization? Not really. The work previously done by the weighted priority diagonalization is here performed by the measure-theoretic density diagonalization in the proof of Theorem 3.13. This result is then used, via Corollary 3.16, to infer Corollary 4.10 from Corollary 4.9. Thus we have not really removed the weighted priority diagonalization. We have, however, clarified its role. It is used only to infer existence from abundance.

If we let $\Delta = \text{all}$, then the observation preceding Corollary 4.9 gives the following well-known result for $K(x|\cdot)$, the conditional Kolmogorov complexity with unbounded resources (i.e., $K(x|\cdot) = \mathrm{KT}^{\infty}(x|\cdot) = \mathrm{KS}^{\infty}(x|\cdot)$).

COROLLARY 4.11 (Martin-Löf [29]). *If* $f: \mathbf{N} \to \mathbf{N}$ *and* $\sum_{n=0}^{\infty} 2^{-f(n)}$ *converges, then a measure* 1 *set of the sequences* $x \in \{0,1\}^{\infty}$ *have* $K(x|n) > n - f(n)$ *a.e.*

Although Corollaries 4.9 and 4.11 are presented here as consequences of Theorem 4.5, it is important to remember that Corollary 4.11 was historically the first such result.

The lower bounds we have given for space-bounded Kolmogorov complexity are fairly tight in the simple sense that they are not too far from the upper bounds given by (4.4) and (4.6). In fact, Martin-Löf [29] showed that the almost everywhere lower bound given by Corollary 4.11 is tight in the much stronger sense that if $f: \mathbf{N} \to \mathbf{N}$ is computable and $\sum_{n=0}^{\infty} 2^{-f(n)}$ diverges, then *every* binary sequence x has $K(x|n) < n - f(n)$ i.o. Thus the convergence/divergence behavior of $\sum_{n=0}^{\infty} 2^{-f(n)}$ determines whether f grows quickly enough that $K(x|n)$ can (and usually does) eventually stay above $n - f(n)$. In the following theorem we modify Martin-Löf's argument to give an infinitely often upper bound on space-bounded conditional program size. This shows that the almost everywhere lower bound given by Theorem 4.5 is very tight. (Ko [17] has proven a similar result.)

THEOREM 4.12. *If* $f: \mathbf{N} \to \mathbf{N}$ *is such that* $f \in$ *pspace and* $\sum_{n=0}^{\infty} 2^{-f(n)}$ *diverges, then there is a polynomial* q *such that every binary sequence* $x \in \{0,1\}^{\infty}$ *has* $KS^{q}(x|n) < n - f(n)$ *i.o.*

Proof. Let $g: \mathbf{N} \to \mathbf{N}$ be computed by the following algorithm.

```
begin
    input n;
    r, s, t := 0, 0, 0;
    for i := 0 to n − 1 do
    begin
        if t ≥ s then r, s := r + 1, 2(s + 2^{r − f(r)});
        t := t + 2^{−f(i)}
    end for-loop;
    output r
end g.
```

It is clear that $g \in \text{pspace}$. We show that g is nondecreasing and onto with

$$\sum_{n=0}^{\infty} 2^{-f(n)-g(n)} = \infty. \tag{4.11}$$

By inspection and induction, the following conditions hold at the beginning of cycle i of the for-loop.

$$r = g(i) \tag{4.12}$$

$$s = 2^r \sum_{j=0}^{r-1} 2^{-f(j)} \tag{4.13}$$

$$t = \sum_{j=0}^{i-1} 2^{-f(j)}. \tag{4.14}$$

It follows that g is nondecreasing with $g(0) = 0$ and range closed downward, i.e., $r_1 \leqslant r_2 \in \text{range}(g)$ implies $r_1 \in \text{range}(g)$. Since $\sum_{n=0}^{\infty} 2^{-f(n)} = \infty$, it follows by (4.14) that g is onto.

Now choose n such that $g(n+1) = g(n) + 1$. Then $t \geqslant s$ in cycle n of the for-loop in the computation of $g(n+1)$. By (4.12–4.14) this implies that

$$\sum_{j=0}^{n-1} 2^{-f(j)} \geqslant 2^{g(n)} \sum_{j=0}^{g(n)-1} 2^{-f(j)},$$

whence

$$\sum_{j=0}^{n} 2^{-f(j)-g(j)} \geqslant 2^{-g(n)} \sum_{j=0}^{n} 2^{-f(j)} \geqslant \sum_{j=0}^{g(n)-1} 2^{-f(j)}. \tag{4.15}$$

Since g is nondecreasing and unbounded and $\sum_{n=0}^{\infty} 2^{-f(n)} = \infty$, (4.11) follows from (4.15). Thus g has the desired properties.

For each $w \in \{0,1\}^*$, define a sequence $w_0, w_1, w_2, \ldots$ of strings $w_i \in \{0,1\}^{|w|}$ by the recursion

$$w_0 = w,$$

$$w_{i+1} = \begin{cases} \text{next}(w_i) & \text{if } w_i \notin \{1\}^* \\ 0^{|w_i|} & \text{if } w_i \in \{1\}^*. \end{cases}$$

This construction has the easily verified property that, for all $w \in \{0,1\}^*$ and $j \in \mathbf{N}$,

$$\{0,1\}^{|w|} = \{w_i \mid j \leqslant i < j + 2^{|w|}\}. \tag{4.16}$$

Now define $F: \mathbf{N}^2 \to \{0,1\}^*$ by

$$F(t,n) = \begin{cases} \lambda & \text{if } n = 0 \text{ or } t > h(n) \\ (F(h(n-1), n-1)1)_t & \text{otherwise,} \end{cases}$$

where $h(n) = \max\{0, 2^{n-f(n)-g(n)} - 1\}$. Note that

$$F(0, n+1) = F(h(n), n)1 \tag{4.17}$$

for all $n \in \mathbf{N}$. We are primarily interested in the strings $F(t, n)$ for $1 \leqslant t \leqslant h(n)$. For each $n \in \mathbf{N}$, these strings form an "interval" of lexicographically successive strings in $\{0,1\}^n$, possibly "wrapping around" from 1^n to 0^n. For each $m, n \in \mathbf{N}$ with $m \geqslant n$, let B_n^m be the set of all strings $w \in \{0,1\}^m$ such that $F(t, n) \sqsubseteq w$ for some $1 \leqslant t \leqslant h(n)$. Note that $|B_n^m| = 2^{m-n}|B_n^n| = 2^{m-n}h(n)$.

Let $n \in \mathbf{N}$ be arbitrary for a moment. By (4.11) there exists $m \geqslant n$ such that

$$\sum_{k=n}^{m} 2^{-f(k)-g(k)} \geqslant 3.$$

Then

$$\begin{aligned}
\sum_{k=n}^{m} |B_k^m| &= \sum_{k=n}^{m} 2^{m-k}h(k) \\
&\geqslant \sum_{k=n}^{m} 2^{m-k}(2^{k-f(k)-g(k)} - 1) \\
&= 2^m \sum_{k=n}^{m} (2^{-f(k)-g(k)} - 2^{-k}) \\
&\geqslant 2^m \left[\sum_{k=n}^{m} 2^{-f(k)-g(k)} - \sum_{k=0}^{\infty} 2^{-k} \right] \\
&\geqslant 2^m(3-2) = 2^m.
\end{aligned}$$

It follows easily by (4.16) and (4.17) that $\bigcup_{k=n}^{m} B_k^m = \{0,1\}^m$. This argument shows that, for every $n \in \mathbf{N}$ and $x \in \{0,1\}^\infty$, there exist $k \geqslant n$ and $1 \leqslant t \leqslant h(k)$ such that $F(t, k) \sqsubseteq x$. That is, for every $x \in \{0,1\}^\infty$, there exist infinitely many $n \in \mathbf{N}$ such that $F(t, n) \sqsubseteq x$ for some $1 \leqslant t \leqslant 2^{n-f(n)-g(n)} - 1$.

Since $f, g \in$ pspace, there is a machine M that, given inputs t, n in binary, outputs $F(t, n)$ in space polynomial in n. It follows by the preceding paragraph that there is a polynomial q' such that

$$\mathrm{KS}_M^{q'}(x \mid n) \leqslant n - f(n) - g(n) \text{ i.o.}$$

for all $x \in \{0,1\}^\infty$. It follows by (4.2) that there exist a polynomial q and a constant c such that

$$\mathrm{KS}^q(x \mid n) \leqslant n - f(n) - g(n) + c \text{ i.o.}$$

for all $x \in \{0,1\}^\infty$. Since g is nondecreasing and unbounded, this proves the theorem. ∎

COROLLARY 4.13. *There is a polynomial q such that every binary sequence $x \in \{0,1\}^\infty$ has $KS^q(x|n) < n - \log n$ i.o.*

In ESPACE, we still have a significant gap between the $n - n^\varepsilon$ lower bound of Corollary 4.6 and the $n - \log n$ upper bound of Corollary 4.13. The following result, due to David Juedes, shows that the $n - n^\varepsilon$ lower bound is tight in ESPACE.

THEOREM 4.14 (Juedes [14]). *Let $q(n) = n^2$. For every $x \in ESPACE$, there exists $\varepsilon > 0$ such that $KS^q(x|n) < n - n^\varepsilon$ a.e.*

Note that the series $\sum_{n=1}^{\infty} 2^{-n^\varepsilon}$ is convergent (in fact, p-convergent), so the upper bound of Theorem 4.14 is tighter than the more general bound of Theorem 4.12.

We now give almost everywhere lower bounds for time-bounded Kolmogorov complexity in uniform time complexity classes.

THEOREM 4.15. *Suppose that $i \in \mathbf{N}$, $g \in G_i$, and $\sigma \in p_{i+1}$ is a selector such that the series $\sum_{n=0}^{\infty} 2^{-\#\sigma(n)\alpha}$ is p_{i+1}-convergent for some real $\alpha < 1$. Then for every $q \in G_{i+1}$, the set of all $x \in \{0,1\}^\infty$ such that $KT^q(x \wedge \sigma|n) > g(\log \#\sigma(n))$ a.e. has p_{i+1}-measure 1, hence measure 1 in E_{i+1}.*

Proof. We follow the proof of Theorem 4.4. In the definitions of X_n and B_n, replace KS by KT, $\#\sigma(n) - f(\#\sigma(n))$ by $g(\log \#\sigma(n))$, and $q(n)$ space by $q(n)$ time. Then $|B_n| < 2^{g(\log \#\sigma(n))}$ is in $2^{G_i(\log G_{i+1})} = 2^{G_i(G_i(\log n))} = 2^{G_i(\log n)} = G_{i+1}(n)$, so $d \in \mathrm{p}_{i+1}$ by (4.9) and (4.10). As in Theorem 4.4, d is a 1-DS and each d_n covers X_n. Finally,

$$\begin{aligned} d_n(\lambda) &= \sum_{\pi \in B_n} P(Z_{n,\pi}) = 2^{-\#\sigma(n)} |B_n| \\ &< 2^{g(\log \#\sigma(n)) + 1 - \#\sigma(n)} < 2^{-\#\sigma(n)\alpha} \end{aligned}$$

for all sufficiently large n, so $\sum_{n=0}^{\infty} d_n(\lambda)$ is p_{i+1}-convergent. ∎

The $g(\log \#\sigma(n))$ lower bound of Theorem 4.15 is asymptotically much smaller than the $\#\sigma(n) - f(\#\sigma(n))$ lower bound of Theorem 4.4. More importantly, the magnitude of the $g(\log \#\sigma(n))$ lower bound in Theorem 4.15 varies directly with the time bound of the uniform complexity class: greater values of i yield greater lower bounds in E_i. Is this relationship an actual property of time complexity classes, or is it merely an artifact of an inadequate analysis? This is a crucial open question that will probably be difficult to answer.

The following almost everywhere lower bound on time-bounded conditional program size is an immediate consequence of Theorem 4.15.

THEOREM 4.16. *If $i \in \mathbf{N}$, $g \in G_i$, and $q \in G_{i+1}$, then the set of all $x \in \{0,1\}^\infty$ such that $KT^q(x|n) > g(\log n)$ a.e. has p_{i+1}-measure 1, hence measure 1 in E_{i+1}.*

The cases $i = 1, 2$ of Theorem 4.16 give polylogarithmic and superpolylogarithmic lower bounds on KT-complexity almost everywhere in E_2 and E_3, respectively.

5. Circuit-Size Complexity

We now use Theorems 4.4 and 4.15 to derive almost everywhere lower bounds on the Boolean circuit-size complexity of binary sequences in exponential complexity classes.

Our circuit terminology is standard. We define a (*Boolean*) *circuit* to be a directed acyclic graph γ with vertex set $I \cup G$, where $I = \{w_1, ..., w_n\}$ is the set of *inputs* ($n \geqslant 0$) and $G = \{g_1, ..., g_s\}$ is the set of *gates* ($s \geqslant 1$). Each input has indegree 0 and each gate has indegree 0, 1, or 2. Each gate of indegree 0 is labeled either by the constant 0 or by the constant 1. Each gate of indegree 1 is labeled either by the identity function ID: $\{0,1\} \to \{0,1\}$ or by the negation function NOT: $\{0,1\} \to \{0,1\}$. Each gate of indegree 2 is labeled either by the conjunction AND: $\{0,1\}^2 \to \{0,1\}$ or by the disjunction OR: $\{0,1\}^2 \to \{0,1\}$. The *output gate* g_s has outdegree 0. The other gates and the inputs have unrestricted outdegree. The *size* of such a circuit γ is $\text{size}(\gamma) = |G| = s$, the number of gates.

An n-input circuit γ *computes* a Boolean function γ: $\{0,1\}^n \to \{0,1\}$ in the usual way. For $w \in \{0,1\}^n$, $\gamma(w)$ is the value computed at the output gate g_s when the inputs are assigned the bits $w_1, ..., w_n$ of w. The *set computed by* an n-input circuit γ is then the set of all $w \in \{0,1\}^n$ such that $\gamma(w) = 1$.

It will be convenient to abbreviate

$$x[\text{length } n] = x[2^n - 1 .. 2^{n+1} - 2]$$

for $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$. We will also define the *graph* of an n-input circuit γ to be the 2^n-bit string

$$\text{graph}(\gamma) = \gamma(s_{2^n-1}) \cdots \gamma(s_{2^{n+1}-2}),$$

where $s_{2^n-1}, ..., s_{2^{n+1}-2}$ are the successive strings of length n. Thus, if x is the characteristic sequence of a language L, then γ computes $L \cap \{0,1\}^n$ if and only if $\text{graph}(\gamma) = x[\text{length } n]$.

By well-known techniques we fix a one-to-one *coding scheme*

$$\Theta\text{: } \{\text{circuits}\} \to \{0,1\}^*,$$

a (small) constant $k_\Theta \in \mathbf{N}$, and a polynomial-time computable *circuit interpreter*

$$I_\Theta\text{: } \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^{\leqslant 1}$$

with the following properties.

(i) For each n-input circuit γ, $|\Theta(\gamma)| \leqslant k_\Theta \, \text{size}(\gamma) \log[n + \text{size}(\gamma)]$.

(ii) If γ_1 and γ_2 are n-input circuits with $\text{size}(\gamma_1) < \text{size}(\gamma_2)$, then $\Theta(\gamma_1)$ lexicographically precedes $\Theta(\gamma_2)$.

(iii) If $y = \Theta(\gamma)$, where γ is a $|w|$-input circuit, then $I_\Theta(y, w) = \gamma(w)$.

(iv) If there is no $|w|$-input circuit γ such that $y = \Theta(\gamma)$, then $I_\Theta(y, w) = \lambda$.

An n-input *circuit code* is a binary string $\Theta(\gamma)$, where γ is an n-input circuit. We sometimes write size($\Theta(\gamma)$) for size(γ) and graph($\Theta(\gamma)$) for graph(γ).

The *circuit-size complexity of a language* $L \subseteq \{0,1\}^*$ is the function $CS_L: \mathbf{N} \to \mathbf{N}$ defined by

$$CS_L(n) = \min\{\text{size}(\gamma) \mid \gamma \text{ computes } L \cap \{0,1\}^n\}.$$

The *circuit-size complexity of a binary sequence* $x \in \{0,1\}^\infty$ is the function $CS_x: \mathbf{N} \to \mathbf{N}$ defined by

$$CS_x(n) = \min\{\text{size}(\gamma) \mid \text{graph}(\gamma) = x[\text{length } n]\}.$$

Note that this is precisely the circuit-size complexity of the language whose characteristic sequence is x.

For each function $f: \mathbf{N} \to \mathbf{N}$ we define the circuit-size complexity classes

$$\text{SIZE}(f) = \{x \in \{0,1\}^\infty \mid CS_x(n) \leqslant f(n) \text{ a.e.}\},$$

$$\text{SIZE}^{\text{i.o.}}(f) = \{x \in \{0,1\}^\infty \mid CS_x(n) \leqslant f(n) \text{ i.o.}\}.$$

For a set C of functions from $\mathbf{N}$ to $\mathbf{N}$ we then define the classes

$$\text{SIZE}(C) = \bigcup_{f \in C} \text{SIZE}(f),$$

$$\text{SIZE}^{\text{i.o.}}(C) = \bigcup_{f \in C} \text{SIZE}^{\text{i.o.}}(f).$$

Identifying languages with their characteristic sequences, $\text{SIZE}(G_0)$ is the set of all languages having linear-size circuits and $\text{SIZE}(G_1)$ is the set of all languages having polynomial-size circuits. Following standard usage, we write P/Poly for $\text{SIZE}(G_1)$. We also write $\text{P/Poly}^{\text{i.o.}}$ for $\text{SIZE}^{\text{i.o.}}(G_1)$.

Notation 5.1. Throughout this section we work with the selector $\hat{\sigma}$ defined by

$$\hat{\sigma}(n) = \mathsf{T}^{\lfloor n/2 \rfloor} \perp^{\lceil n/2 \rceil}.$$

In the terminology of Section 4, the selector $\hat{\sigma}$ requires a program to correctly decide the last $\#\hat{\sigma}(n) = \lceil n/2 \rceil$ bits of an n-bit prefix $x[0..n-1]$. In particular, $\hat{\sigma}(2^{n+1}-1)$ requires a program to correctly decide the substring $x[\text{length } n]$ of x.

Our derivation of circuit-size lower bounds from space-bounded selective Kolmogorov complexity lower bounds employs the following relationship.

LEMMA 5.2. *There exist a polynomial q and a constant $\hat{c}$ such that, for every binary sequence $x \in \{0,1\}^\infty$ and every $n \in \mathbf{N}$,*

$$KS^q(x \wedge \hat{\sigma} \mid 2^{n+1}-1) \leqslant g_x(n)[\hat{c} + \log g_x(n)],$$

where $g_x(n) = \max\{n, CS_x(n)\}$. (This refines a result of Abu-Mostafa [46].)

Proof. Call a string $y \in \{0,1\}^\infty$ *novel* for n if y is an n-input circuit code and, for every n-input circuit code y' that lexicographically precedes y, $\text{graph}(y') \neq \text{graph}(y)$. The predicate "$y$ is novel for n" can easily be tested in space that is polynomial in $n + |y|$. Let $y_1, \ldots, y_{J(n)}$ be the lexicographic enumeration of those strings that are novel for n. It is routine to design a machine M that takes inputs $t, N \in \mathbf{N}$ in binary and has the following property. If $N = 2^{n+1} - 1$ and $1 \leqslant t \leqslant J(n)$, then $M(t, N) = 0^{2^n - 1}\,\text{graph}(y_t)$, and this computation is carried out in space that is polynomial in N. It follows by (4.2) that there exist a polynomial q and a constant c such that

$$\text{KS}^q(x \wedge \hat{\sigma} \mid 2^{n+1} - 1) \leqslant c + |t| \tag{5.1}$$

whenever $x[\text{length } n] = \text{graph}(y_t)$ for some $1 \leqslant t \leqslant J(n)$.

We now estimate the number $H_x(n)$ of strings y that are novel for n and have $\text{size}(y) \leqslant g_x(n)$. (Such an estimate was first computed by Shannon [38]. Minor variations of Shannon's estimate have appeared many times. The argument here, included for completeness, is similar to that of Balcázar, Díaz, and Gabarró [4].) In an n-input circuit with s gates, each gate has fewer than $6(n+s)^2$ possible specifications of its function and the sources of its inputs. Thus there are fewer than $6^s(n+s)^{2s}$ such circuits. Each of these circuits is functionally equivalent to the $(s-1)!$ circuits obtained by permuting its $s-1$ nonoutput gates (and adjusting the inputs to the output gate accordingly), so the number of functionally distinct such circuits is less than $6^s(n+s)^{2s}/(s-1)! = s6^s(n+s)^{2s}/s!$. This is less than $[12(n+s)^2]^s/s!$. Using the weak Stirling approximation $s! > (s/e)^s$, then, the number of distinct such circuits is less than $[12e(n+s)^2/s]^s$. Since $g_x(n) \geqslant n$ and every circuit with fewer than $g_x(n)$ gates can be simulated by a circuit with exactly $g_x(n)$ gates, it follows that

$$H_x(n) < \left[\frac{12e(n + g_x(n))^2}{g_x(n)}\right]^{g_x(n)} \leqslant [48e g_x(n)]^{g_x(n)} \tag{5.2}$$

for all $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$.

By the monotonicity of the circuit coding Θ, for every $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$, there is some $1 \leqslant t \leqslant H_x(n)$ such that $x[\text{length } n] = \text{graph}(y_t)$. Setting

$$\hat{c} = 1 + c + \log(48e),$$

it follows from (5.1) and (5.2) that

$$\begin{aligned} \text{KS}^q(x \wedge \hat{\sigma} \mid 2^{n+1} - 1) &\leqslant c + |t| \\ &\leqslant c + 1 + \log H_x(n) \\ &\leqslant c + 1 + g_x(n) \log[48e g_x(n)] \\ &\leqslant g_x(n)[\hat{c} + \log g_x(n)] \end{aligned}$$

for all $x \in \{0,1\}_\infty$ and $n \in \mathbf{N}$. ∎

Our almost everywhere lower bound for circuit size in ESPACE can now be derived from Theorem 4.4.

THEOREM 5.3. *For every* $\alpha < 1$, *the set of all* $x \in \{0,1\}^\infty$ *such that* $CS_x(n) > (2^n/n)(1+\alpha \log n/n)$ *a.e. has pspace-measure* 1, *hence measure* 1 *in ESPACE.*

Proof. Fix $0 < \alpha < 1$ and write $\beta = 1 + \alpha \log n/n$ for convenience. Assume for a moment that

$$CS_x(n) \leqslant \frac{2^n}{n}\left(1 + \frac{\alpha \log n}{n}\right) \text{ i.o.} \tag{5.3}$$

Choosing q and $\hat{c}$ as in Lemma 5.2, we then have

$$\begin{aligned} KS^q(x \wedge \hat{\sigma} \mid 2^{n+1}-1) &\leqslant \frac{2^n}{n}\beta\left[\hat{c} + \log\left(\frac{2^n}{n}\beta\right)\right] \\ &= 2^n - \frac{2^n}{n}\left[(\beta-\alpha)\log n - \beta(\hat{c} + \log\beta)\right] \\ &\leqslant 2^n - \frac{2^n}{n}\left[(1-\alpha)\log n - \beta(\hat{c} + \log\beta)\right] \text{ i.o.} \end{aligned}$$

Since $\beta(\hat{c} + \log\beta) \to \hat{c}$ as $n \to \infty$, it follows that

$$KS^q(x \wedge \hat{\sigma} \mid 2^{n+1}-1) \leqslant 2^n - \frac{2^n}{n}\left[(1-\alpha)\log n - 2\hat{c}\right] \text{ i.o.}$$

Rewriting this with the change of variable $N = 2^{n+1}-1$ gives

$$KS^q(x \wedge \hat{\sigma} \mid N) \leqslant \#\hat{\sigma}(N) - f(\#\hat{\sigma}(N)) \text{ i.o.}, \tag{5.4}$$

where $f(k) = (k/\log k)[(1-\alpha)\log\log k - 2\hat{c}]$.

Now fix a constant $k_0 \in \mathbf{N}$ such that

$$f(k) \geqslant \sqrt{k}\log e \qquad \text{and} \qquad \left(\frac{e}{2}\right)^{\sqrt{k}} \geqslant 4(k+1)$$

hold whenever $k \geqslant k_0$. Set $g(j) = 2(j^2 + k_0 + 1)$ for all $j \in \mathbf{N}$. Then $g \in \mathrm{p}$ and

$$\begin{aligned} \sum_{n=g(j)}^{\infty} 2^{-f(\#\hat{\sigma}(n))} &= \sum_{n=g(j)}^{\infty} 2^{-f(\lceil n/2 \rceil)} = 2 \sum_{n=j^2+k_0+1}^{\infty} 2^{-f(n)} \\ &\leqslant 2 \sum_{n=j^2+k_0+1}^{\infty} e^{-\sqrt{n}} \leqslant 2\int_{j^2+k_0}^{\infty} e^{-\sqrt{t}}\, dt \\ &= 4e^{-\sqrt{j^2+k_0}}\left(1 + \sqrt{j^2+k_0}\right) \leqslant 2^{-j} \end{aligned}$$

for all $j \in \mathbf{N}$; i.e., g testifies that the series $\sum_{n=0}^{\infty} 2^{-f(\#\hat{\sigma}(n))}$ is p-convergent. It follows by Theorem 4.4 that the set of all $x \in \{0,1\}^{\infty}$ satisfying (5.4) has pspace-measure 0. Since (5.3) implies (5.4), this proves the theorem. ∎

As an immediate consequence of Theorem 5.3, we have the following strengthening of Shannon's almost everywhere lower bound [38] on circuit size.

THEOREM 5.4. *For every real number* $\alpha < 1$, *almost every binary sequence* $x \in \{0,1\}^{\infty}$ *has circuit-size complexity* $CS_x(n) > (2^n/n)(1 + \alpha \log n/n)$ *a.e.*

The distribution of complexities between this lower bound and the $(2^n/n)(1 + O(1/\sqrt{n}))$ upper bound of Lupanov [23] remains an open question.

Theorem 5.3 extends the following known result by increasing the lower bound and by substituting "a.e." for "i.o."

COROLLARY 5.5 (Lutz [24]). *If* $f: \mathbf{N} \to \mathbf{N}$ *is such that* $f \in pspace$ *and* $f(n) = o(2^n/n)$, *then the set of binary sequences* $x \in \{0,1\}^{\infty}$ *such that* $CS_x(n) > f(n)$ *i.o. has pspace-measure* 1, *hence measure* 1 *in ESPACE.*

COROLLARY 5.6 (Lutz [24]). $\mu(P/Poly \mid ESPACE) = 0$.

In fact, we now have a stronger result.

COROLLARY 5.7. $\mu(P/Poly^{i.o.} \mid ESPACE) = 0$.

The following consequence of Theorem 5.3 (via Corollary 5.7) was in the fact the starting point for research leading to Theorem 5.3.

COROLLARY 5.8 (Kannan [15]). $ESPACE \not\subseteq P/Poly^{i.o.}$.

We now consider circuit-size complexity in uniform time complexity classes. For this we use the following relationship between circuit size and time-bounded selective Kolmogorov complexity.

LEMMA 5.9. *There exist a polynomial* q *and constants* $\hat{c}_1$ *and* $\hat{c}_2$ *such that, for every binary sequence* $x \in \{0,1\}^{\infty}$ *and every* $n \in \mathbf{N}$,

$$KT^q(x \wedge \hat{\sigma} \mid 2^{n+1} - 1) \leqslant \hat{c}_1 g_x(n) \log g_x(n) + \hat{c}_2,$$

where $g_x(n) = \max\{n, CS_x(x)\}$. (*This refines a result of Abu-Mostafa* [46].)

Proof. Using the circuit interpreter I_{Θ} we can design a machine M such that if $N = 2^{n+1} - 1$ and $y = \Theta(\gamma)$, where γ is an n-input circuit, then

$$M(y, N) = 0^{2^n - 1}\,\mathrm{graph}(\gamma)$$

in $\leqslant q'(N + \mathrm{size}(\gamma))$ time, where q' is a polynomial. In fact, by the Lupanov upper

bound there is a polynomial q'' such that for every string $z \in \{0,1\}^{2^n}$ there is a circuit code y such that

$$M(y, N) = 0^{2^n - 1}z$$

in $\leqslant q''(N)$ time. It follows by our choice of circuit coding scheme that

$$\begin{aligned} \mathrm{KT}^{q''}_M(x \wedge \hat{\sigma} \mid 2^{n+1} - 1) &\leqslant k_\Theta \, \mathrm{CS}_x(n) \log[n + \mathrm{CS}_x(n)] \\ &\leqslant \hat{c}_1 \, g_x(n) \log g_x(n) \end{aligned}$$

for every $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$, where $\hat{c}_1 = 2k_\Theta$. By (4.1), then, there exist a polynomial q and a constant $\hat{c}_2$ such that

$$\mathrm{KT}^q(x \wedge \hat{\sigma} \mid 2^{n+1} - 1) \leqslant \hat{c}_1 \, g_x(n) \log g_x(n) + \hat{c}_2$$

for all $x \in \{0,1\}^\infty$ and $n \in \mathbf{N}$. ∎

Almost everywhere lower bounds for circuit size are now easily derived from Theorem 4.15.

THEOREM 5.10. *If $i \geqslant 1$ and $f \in G_i$, then the set of all $x \in \{0,1\}^\infty$ such that $CS_x(n) > f(n)$ a.e. has p_{i+1}-measure 1, hence measure 1 in E_{i+1}.*

Proof. If $x \in \{0,1\}^\infty$ is such that

$$\mathrm{CS}_x(n) \leqslant f(n) \text{ i.o.},$$

then Lemma 5.9 tells us that there exist functions $g, q \in G_i$ such that

$$\mathrm{KT}^q(x \wedge \hat{\sigma} \mid n) \leqslant g(\log \#\hat{\sigma}(n)) \text{ i.o.} \tag{5.5}$$

Since the set of all $x \in \{0,1\}^\infty$ satisfying (5.5) has p_{i+1}-measure 0 by Theorem 4.15, the present theorem follows. ∎

COROLLARY 5.11. *For $i \in \mathbf{N}$, $SIZE^{i.o.}(G_i)$ has p_{i+2}-measure 0, hence measure 0 in E_{i+2}.*

Proof. Since $\mathrm{SIZE}^{\mathrm{i.o.}}(G_i) \subseteq \mathrm{SIZE}^{\mathrm{i.o.}}(\hat{g}_{i+1})$ and $\hat{g}_{i+1} \in G_{i+1}$, this follows immediately from Theorem 5.10. ∎

COROLLARY 5.12. *$P/Poly^{i.o.}$ has p_3-measure 0, so $\mu(P/Poly^{\mathrm{i.o.}} \mid E_3) = 0$.*

COROLLARY 5.13. *For each $k \in \mathbf{N}$, $SIZE^{i.o.}(n^k)$ has p_2-measure 0, so $\mu(SIZE^{i.o.}(n^k) \mid E_2) = 0$.*

Theorem 5.10 extends a result of [24] by substituting "a.e." for "i.o." Corollaries 5.11, 5.12, and 5.13 then extend results of [24] in like fashion.

Since Wilson [44] has exhibited oracles relative to which $\mathrm{E}_2 \subseteq \mathrm{P/Poly}$ and

$E \subseteq SIZE(G_0)$, Corollaries 5.12 and 5.13 appear to be the strongest results that we can obtain from relativizable techniques.

6. Pseudorandom Sequences

The results of the preceding two sections can now be used to prove lower bounds on the nonuniform complexity of pseudorandom sequences. We first define the measure-theoretic notion of pseudorandomness.

Definition 6.1. A *Δ-test* is a set $X \subseteq \{0,1\}^\infty$ such that $\mu_\Delta(X) = 1$. A binary sequence $x \in \{0,1\}^\infty$ *passes* a Δ-test X if $x \in X$. A binary sequence $x \in \{0,1\}^\infty$ is *Δ-random*, and we write $x \in \mathrm{RAND}(\Delta)$, if x passes all Δ-tests. That is,

$$\mathrm{RAND}(\Delta) = \bigcap \{X \mid \mu_\Delta(X) = 1\}.$$

It is an essential feature of Δ-randomness that it (like the algorithmic randomness of Martin-Löf [28] and the weak randomness of Schnorr [34–37]) is definable in measure-theoretic terms. However, Δ-randomness admits other characterizations, just one of which we mention here. (This follows immediately from Theorem 3.19 and Definition 6.1.)

Theorem 6.2. *A sequence $x \in \{0,1\}^\infty$ is Δ-random if and only if there is no Δ-computable martingale that succeeds on x.*

If $\Delta = \mathrm{rec}$, then Theorem 6.2 tells us that rec-randomness is equivalent to the *martingale randomness* mentioned by van Lambalgen [42, pp. 77–78]. Thus if we let RAND be the set of all algorithmically random sequences of Martin-Löf [28] and RAND_W be the set of all weakly random sequences of Schnorr [36] (see also [41, 42]), then

$$\mathrm{RAND} \subseteq \mathrm{RAND}(\mathrm{rec}) \subseteq \mathrm{RAND}_W.$$

If Δ is a time- or space-bounded complexity class, then Δ-randomness is a notion of pseudorandomness that is at least as strong as (and, we conjecture, stronger than) the time- and space-bounded versions of RAND_W investigated by Schnorr [34, 36]. In any case, such classes $\mathrm{RAND}(\Delta)$ have the following abundance property, which can be regarded as a weak analogue of the existence of a universal test for algorithmic randomness [28]. (See [27] for a proof and further discussion of pseudorandom sequences.)

Theorem 6.3 (Abundance Theorem). *For $i \geqslant 1$, $RAND(p_i)$ is a p_{i+1}-test and $RAND(p_i space)$ is a $p_{i+1} space$-test. That is,*

$$\mu_{p_{i+1}}(RAND(p_i)) = \mu_{p_{i+1} space}(RAND(p_i space)) = 1.$$

Thus $\mu(RAND(p_i) \mid E_{i+1}) = \mu(RAND(p_i space) \mid E_{i+1} SPACE) = 1$.

Thus almost every sequence in E_{i+1} is p_i-random and almost every sequence in $E_{i+1}SPACE$ is $p_i$space-random. That is, Definition 6.1 is sufficiently weak to provide an abundance of deterministically computed pseudorandom sequences. On the other hand, every singleton subset of $R(\Delta)$ has Δ-measure 0, so no sequence in E_i is p_i-random and no sequence in E_iSPACE is $p_i$space-random. Thus we immediately have lower bounds on the uniform complexities of pseudorandom sequences. The following results give lower bounds on the nonuniform complexities of pseudorandom sequences.

THEOREM 6.4. *If $x \in \{0,1\}^\infty$ is pspace-random, then*

$$KS^q(x \mid n) > n - f(n) \text{ a.e.} \tag{6.1}$$

for every polynomial q and every $f \in$ pspace such that $\sum_{n=0}^{\infty} 2^{-f(n)}$ is p-convergent; and

$$CS_x(n) > \frac{2^n}{n}\left(1 + \frac{\alpha \log n}{n}\right) \text{ a.e.} \tag{6.2}$$

for every real number $\alpha < 1$.

Proof. By Theorems 4.5 and 5.3, conditions (6.1) and (6.2) are pspace-tests. ∎

THEOREM 6.5. *If $x \in \{0,1\}^\infty$ is p_i-random, where $i \geqslant 1$, then*

$$KT^q(x \mid n) > g(\log n) \text{ a.e.} \tag{6.3}$$

for all $g \in G_{i-1}$ and $q \in G_i$. If $i \geqslant 2$, then we also have

$$CS_x(n) > f(n) \text{ a.e.} \tag{6.4}$$

for all $f \in G_{i-1}$.

Proof. By Theorems 4.16 and 5.10, conditions (6.3) and (6.4) are p_i-tests. ∎

COROLLARY 6.6. $RAND(p_2) \cap P/Poly^{i.o.} = \varnothing$.

That is, every p_2-random sequence has superpolynomial circuit-size complexity almost everywhere.

7. CONCLUSION

We have proven several results of the following general form.

> *Almost every* problem in the uniform complexity class $\mathscr{C}$ has very high nonuniform complexity *almost everywhere.*

For $\mathscr{C} =$ ESPACE, these results give strong instances of the Shannon effect.

For time-bounded classes $\mathscr{C}$, the results are distributionally strong but leave

$E \not\subseteq P/Poly$ and other important conjectures unresolved. We have, however, shed some structural light on such questions. For example, Theorem 5.3 tells us that at least one of the following is true.

(i) $E \not\subseteq SIZE^{i.o.}((2^n/n)(1 + (\alpha \log n/n)))$ for every real $\alpha < 1$.

(ii) E is a measure 0 subset of ESPACE.

Condition (i) is much stronger than the $E \not\subseteq P/Poly$ conjecture. By the work of Hartmanis and Yesha [11], condition (ii) implies, and is probably stronger than, the conjecture that $P \subsetneq P/Poly \cap PSPACE$.

ACKNOWLEDGMENTS

I thank Yaser Abu-Mostafa, Leonid Levin, Elvira Mayordomo, David Juedes, David Martin, Giora Slutzki, and Josef Breutzmann for helpful discussions and remarks.

REFERENCES

1. E. W. ALLENDER, Some consequences of the existence of pseudorandom generators, *J. Comput. System Sci.* **39** (1989), 101–124.
2. E. W. ALLENDER AND O. WATANABE, Kolmogorov complexity and degrees of tally sets, *Inform. Comput.* **86** (1990), 160–178.
3. J. L. BALCÁZAR AND R. V. BOOK, Sets with small generalized Kolmogorov complexity, *Acta Inform.* **23** (1986), 679–688.
4. J. L. BALCÁZAR, J. DÍAZ, AND J. GABARRÓ, "Structural Complexity I," Springer-Verlag, New York, 1988.
5. M. BLUM AND S. MICALI, How to generate cryptographically strong sequences of pseudo-random bits, *SIAM J. Comput.* **13** (1984), 850–864.
6. G. J. CHAITIN, On the length of programs for computing finite binary sequences, *J. Assoc. Comput. Mach.* **13** (1966), 547–569.
7. R. I. FREIDZON, Families of recursive predicates of measure zero, *J. Soviet Math.* **6** (1976), 449–455.
8. O. GOLDREICH, S. GOLDWASSER, AND S. MICALI, How to construct random functions, *J. Assoc. Comput. Mach.* **33** (1986), 792–807.
9. P. R. HALMOS, "Measure Theory," Springer-Verlag, New York, 1950.
10. J. HARTMANIS, Generalized Kolmogorov complexity and the structure of feasible computations, *in* "Proceedings, 24th IEEE Symposium on the Foundations of Computer Science, 1983," pp. 439–445.
11. J. HARTMANIS AND Y. YESHA, Computation times of NP sets of different densities, *Theoret. Comput. Sci.* **34** (1984), 17–32.
12. D. T. HUYNH, Some observations about the randomness of hard problems, *SIAM J. Comput.* **15** (1986), 1101–1105.
13. D. T. HUYNH, Resource-bounded Kolmogorov complexity of hard languages, *in* "Structure in Complexity Theory," Lecture Notes in Computer Science, Vol. 223, pp. 184–195, 1986.
14. D. W. JUEDES AND J. H. LUTZ, Kolmogorov complexity, complexity cores, and the distribution of hardness, *in* "Kolmogorov Complexity: Theory and Relations to Computational Complexity" (O. Watanabe, Ed.), Springer-Verlag, New York, to appear.
15. R. KANNAN, Circuit-size lower bounds and non-reducibility to sparse sets, *Inform. Control* **55** (1982), 40–56.
16. R. M. KARP AND R. J. LIPTON, Some connections between nonuniform and uniform complexity classes, *in* "Proceedings, 12th ACM Symposium on the Theory of Computing, 1980," pp. 302–309.

17. K. Ko, On the notion of infinite pseudorandom sequences, *Theoret. Comput. Sci.* **48** (1986), 9–33.
18. A. N. Kolmogorov, Three approaches to the quantitative definition of "information," *Probl. Inform. Trans.* **1** (1965), 1–7.
19. A. N. Kolmogorov and V. A. Uspenskii, Algorithms and randomness, *Theory Probab. Appl.* **32** (1987), 389–412.
20. L. A. Levin, One way functions and pseudorandom generators, *Combinatorica* **7** (1987), 357–363.
21. M. Li and P. M. B. Vitanyi, Kolmogorov Complexity and its Applications, *in* "Handbook of Theoretical Computer Science" (J. van Leeuwen, Ed.), Elsevier, Amsterdam/New York, 1990.
22. L. Longpré, "Resource Bounded Kolmogorov Complexity, a Link between Computational Complexity and Information Theory," Ph.D. thesis, Technical Report TR-86-776, Cornell-University, 1986.
23. O. B. Lupanov, On the synthesis of contact networks, *Dokl. Akad. Nauk SSSR* **19** (1958), 23–26.
24. J. H. Lutz, Category and measure in complexity classes, *SIAM J. Comput.* **19** (1990), 1100–1131.
25. J. H. Lutz, Pseudorandom sources for BPP, *J. Comput. System Sci.* **41** (1990), 307–320.
26. J. H. Lutz, Resource-bounded measure, in preparation.
27. J. H. Lutz, Intrinsically pseudorandom sequences, in preparation.
28. P. Martin-Löf, On the definition of random sequences, *Inform. Control* **9** (1966), 602–619.
29. P. Martin-Löf, Complexity oscillations in infinite binary sequences, *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **19** (1971), 225–230.
30. K. Mehlhorn, On the size of sets of computable functions, *in* "Proceedings, 14th IEEE Symposium on the Foundations of Computer Science, 1973," pp. 190–196.
31. K. Mehlhorn, The "almost all" theory of subrecursive degrees is decidable, *in* "Proceedings, 2nd Colloquium on Automata, Languages and Programming," pp. 317–325, Springer Lecture Notes, 1974.
32. A. R. Meyer and E. M. McCreight, Computationally complex and pseudorandom zero-one valued functions, *in* "Theory of Machines and Computations" (Z. Kohavi and A. Paz, Eds.), Academic Press, 1971.
33. J. C. Oxtoby, "Measure and Category," 2nd ed., Springer-Verlag, New York, 1980.
34. C. P. Schnorr, Klassifikation der Zufallsgesetze nach Komplexität und Ordnung, *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **16** (1970), 1–21.
35. C. P. Schnorr, A unified approach to the definition of random sequences, *Math. Systems Theory* **5** (1971), 246–258.
36. C. P. Schnorr, "Zufälligkeit und Wahrscheinlichkeit," Lecture Notes in Mathematics, Vol. 218, Springer-Verlag, Berlin/New York, 1971.
37. C. P. Schnorr, Process complexity and effective random tests, *J. Comput. System Sci.* **7** (1973), 376–388.
38. C. E. Shannon, The synthesis of two-terminal switching circuits, *Bell System Tech. J.* **28** (1949), 59–98.
39. M. Sipser, A complexity-theoretic approach to randomness, *in* "Proceedings, 15th ACM Symposium on the Theory of Computing, 1983," pp. 330–335.
40. R. J. Solomonoff, A formal theory of inductive inference, *Inform. Control* **7** (1964), 1–22, 224–254.
41. V. A. Uspenskii, A. L. Semenov, and A. Kh. Shen', Can an individual sequence of zeros and ones be random? *Russ. Math. Surveys* **45** (1990), 121–189.
42. M. van Lambalgen, "Random Sequences," Ph.D. thesis, Department of Mathematics, University of Amsterdam, 1987.
43. R. E. Wilber, Randomness and the density of hard problems, *in* "Proceedings, 24th IEEE Symposium on the Foundations of Computer Science, 1983," pp. 335–342.
44. C. B. Wilson, Relativized circuit complexity, *J. Comput. System Sci.* **31** (1985), 169–181.
45. A. Yao, Theory and applications of trapdoor functions, *in* "Proceedings, 23rd IEEE Symposium on Foundations of Computer Science, 1982," pp. 80–91.
46. Y. S. Abu-Mostafa, The complexity of information extraction, *IEEE Transactions on Information Theory* **IT-32** (1986), 513–525.